03/12348

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 October 2002 (31.10.2002)

PCT

(10) International Publication Number
WO 02/086071 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/12497

(22) International Filing Date: 19 April 2002 (19.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/285,343 20 April 2001 (20.04.2001) US
60/356,937 14 February 2002 (14.02.2002) US

(71) Applicant *(for all designated States except US)*: **LUDWIG INSTITUTE FOR CANCER RESEARCH** [CH/US]; 605 Third Avenue, New York, NY 10158 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NAKAYAMA, Eiichi** [JP/JP]; Okayama University Medical School, Department of Immunology, 2-5-1, Shikata-cho, Okayama 700-8558 (JP). **ONO, Toshiro** [JP/JP]; Okayama University Medical School, Department of Immunology, 2-5-1, Shikata-cho, Okayama 700-8558 (JP). **OLD, Lloyd, J.** [US/US]; 33rd Floor, 605 Third Avenue, New York, NY 10158 (US).

(74) Agent: **VAN AMSTERDAM, John, R.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

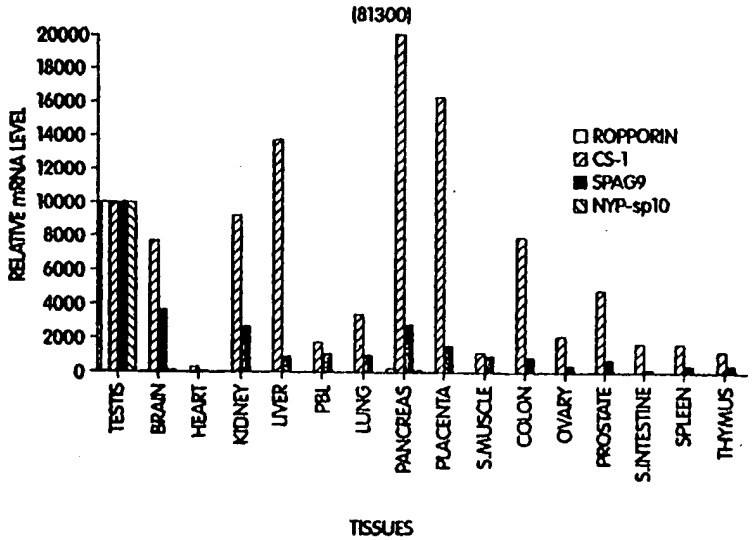
(81) Designated States *(national)*: AU, CA, CN, JP, KR, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: CANCER-TESTIS ANTIGENS




(57) Abstract: CT antigens have been identified by screening known sperm-specific genes for expression in tumors and testis. The invention relates to nucleic acids and encoded polypeptides which are CT antigens expressed in patients afflicted with cancer. The invention provides, inter alia, isolated nucleic acid molecules, expression vectors containing those molecules and host cells transfected with those molecules. The invention also provides isolated proteins and peptides, antibodies to those proteins and peptides and cytotoxic T lymphocytes which recognize the proteins and peptides. Fragments of the foregoing including functional fragments and variants also are provided. Kits containing the foregoing molecules additionally are provided. The molecules provided by the invention can be used in the diagnosis, monitoring, research, or treatment of conditions characterized by the expression of one or more CT antigens.

WO 02/086071 A2

# CANCER-TESTIS ANTIGENS

## Field of the Invention

The invention relates to nucleic acids and encoded polypeptides which are novel cancer-testis antigens expressed in a variety of cancers. The invention also relates to agents which bind the nucleic acids or polypeptides. The nucleic acid molecules, polypeptides coded for by such molecules and peptides derived therefrom, as well as related antibodies and cytolytic T lymphocytes, are useful, *inter alia*, in diagnostic and therapeutic contexts.

## Background of the Invention

It is a little acknowledged fact that the discipline of tumor immunology has been the source of many findings of critical importance in cancer-related as well as cancer-unrelated fields. For example, it was the search for tumor antigens that led to the discovery of the CD8 T cell antigen (1) and the concept of differentiation antigens (2) (and the CD system for classifying cell surface antigens), and to the discovery of p53 (3). The immunogenetic analysis of resistance to viral leukemogenesis provided the first link between the MHC and disease susceptibility (4), and interest in the basis for non-specific immunity to cancer gave rise to the discovery of TNF (5).

Another area of tumor immunology that holds great promise is the category of antigens referred to as cancer/testis (CT) antigens, first recognized as targets for CD8 T cell recognition of autologous human melanoma cells (6, 7). The molecular definition of these antigens was a culmination of prior efforts to establish systems and methodologies for the unambiguous analysis of humoral (8) and cellular (9) immune reactions of patients to autologous tumor cells (autologous typing), and this approach of autologous typing also led to the development of SEREX (serological analysis of cDNA expression libraries) for defining the molecular structure of tumor antigens eliciting a humoral immune response (10).

Although the usefulness of the known CT antigens in the diagnosis and therapy of cancer is accepted, the expression of these antigens in tumors of various types and sources is not universal. Accordingly, there is a need to identify additional CT antigens to provide more targets for diagnosis and therapy of cancer, and for the development of pharmaceuticals useful in diagnostic and therapeutic applications.

## Summary of the Invention

Bioinformatic analysis of sequence databases has been applied to identify sequences having expression characteristics that fit the profile of cancer/testis antigens. Several novel cancer/testis antigens and cancer associated antigens have been identified. The invention provides, *inter alia*, isolated nucleic acid molecules, expression vectors containing those molecules and host cells transfected with those molecules. The invention also provides isolated proteins and peptides, antibodies to those proteins and peptides and CTLs which recognize the proteins and peptides. Fragments and variants of the foregoing also are provided. Kits containing the foregoing molecules additionally are provided. The foregoing can be used in the diagnosis, monitoring, research, or treatment of conditions characterized by the expression of one or more cancer-testis and/or cancer associated antigens.

Prior to the present invention, only a handful of cancer/testis antigens had been identified in the past 20 years. The invention involves the surprising discovery of several sequence clusters (UniGene) in sequence databases that have expression patters that fit the profile of cancer-testis antigens. Other sequence clusters fit the profile of cancer associated antigens. The knowledge that these sequence clusters have these certain expression patterns makes the sequences useful in the diagnosis, monitoring and therapy of a variety of cancers.

The invention involves the use of a single material, a plurality of different materials and even large panels and combinations of materials. For example, a single gene, a single protein encoded by a gene, a single functional fragment thereof, a single antibody thereto, etc. can be used in methods and products of the invention. Likewise, pairs, groups and even panels of these materials and optionally other CT antigen genes and/or gene products can be used for diagnosis, monitoring and therapy. The pairs, groups or panels can involve 2, 3, 4, 5 or more genes, gene products, fragments thereof or agents that recognize such materials. A plurality of such materials are not only useful in monitoring, typing, characterizing and diagnosing cells abnormally expressing such genes, but a plurality of such materials can be used therapeutically. An example of the use of a plurality of such materials for the prevention, delay of onset, amelioration, etc. of cancer cells, which express or will express such genes prophylactically or acutely. Any and all combinations of the genes, gene products, and materials which recognize the genes and gene products can be tested and identified for use according to the invention. It would be far too lengthy to recite all such combinations; those skilled in the art, particularly in view of the teaching contained herein,

will readily be able to determine which combinations are most appropriate for which circumstances.

As will be clear from the following discussion, the invention has *in vivo* and *in vitro* uses, including for therapeutic, diagnostic, monitoring and research purposes. One aspect of the invention is the ability to fingerprint a cell expressing a number of the genes identified according to the invention by, for example, quantifying the expression of such gene products. Such fingerprints will be characteristic, for example, of the stage of the cancer, the type of the cancer, or even the effect in animal models of a therapy on a cancer. Cells also can be screened to determine whether such cells abnormally express the genes identified according to the invention.

According to one aspect of the invention, methods of diagnosing a disorder characterized by expression of a human CT antigen precursor coded for by a nucleic acid molecule are provided. The methods include contacting a biological sample isolated from a subject with an agent that specifically binds to the nucleic acid molecule, an expression product thereof, a fragment of an expression product thereof complexed with an HLA molecule, or an antibody that binds to the expression product, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and determining the interaction between the agent and the nucleic acid molecule or the expression product as a determination of the disorder.

In some embodiments the agent is selected from the group consisting of (a) nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 or a fragment thereof, (b)an antibody that binds to an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, (c)an agent that binds to a complex of an HLA molecule and a fragment of an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and (d) an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 that binds an antibody. Preferred sequences include SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In other embodiments the disorder is characterized by expression of a plurality of human CT antigen precursors and wherein the agent is a plurality of agents, each of which is specific for a different human CT antigen precursor, and wherein said plurality of agents is at

least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8, at least 9 or at least 10 such agents. Preferably the disorder is cancer.

According to another aspect of the invention, methods for determining regression, progression or onset of a condition characterized by expression of abnormal levels of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 are provided. The methods include monitoring a sample, from a patient who has or is suspected of having the condition, for a parameter selected from the group consisting of (i)the protein, (ii)a peptide derived from the protein, (iii) an antibody which selectively binds the protein or peptide, and (iv) cytolytic T cells specific for a complex of the peptide derived from the protein and an MHC molecule, as a determination of regression, progression or onset of said condition. Preferably the sample is assayed for the peptide. Preferred sequences include SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In certain embodiments, the sample is a body fluid, a body effusion, cell or a tissue. In other embodiments, the step of monitoring comprises contacting the sample with a detectable agent selected from the group consisting of (a) an antibody which selectively binds the protein of (i), or the peptide of (ii), (b)a protein or peptide which binds the antibody of (iii), and (c) a cell which presents the complex of the peptide and MHC molecule of (iv). Preferably, the antibody, the protein, the peptide or the cell is labeled with a radioactive label or an enzyme.

In other embodiments, the protein is a plurality of proteins, the parameter is a plurality of parameters, each of the plurality of parameters being specific for a different of the plurality of proteins, at least one of which is a CT antigen protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting SEQ ID NOS:1, 3, 5, 7, 9 and 63. In further embodiments, the protein is a plurality of proteins, at least one of which is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and wherein the parameter is a plurality of parameters, each of the plurality of parameters being specific for a different of the plurality of proteins.

According to a further aspect of the invention, pharmaceutical preparations for a human subject are provided. The pharmaceutical preparations include an agent which when administered to the subject enriches selectively the presence of complexes of an HLA molecule and a human CT antigen peptide, and a pharmaceutically acceptable carrier,

wherein the human CT antigen peptide is a fragment of a human CT antigen encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

In some embodiments, the agent comprises a plurality of agents, each of which enriches selectively in the subject complexes of an HLA molecule and a different human CT antigen peptide, wherein at least one of the human CT antigens is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. Preferably the plurality is at least two, at least three, at least four or at least five different such agents.

In still other embodiments, the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56), or the agent comprises a plurality of agents, at least one of which is a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56), or an expression product thereof, each of which enriches selectively in the subject complexes of an HLA molecule and a different human CT antigen.

In other preferred embodiments, the agent is selected from the group consisting of (1) an isolated polypeptide comprising the human CT antigen peptide, or a functional variant thereof, (2) an isolated nucleic acid operably linked to a promoter for expressing the isolated polypeptide, or functional variant thereof, (3) a host cell expressing the isolated polypeptide, or functional variant thereof, and (4) isolated complexes of the polypeptide, or functional variant thereof, and an HLA molecule.

Preferred pharmaceutical preparations also include an adjuvant.

In still other embodiments, the agent is a cell expressing an isolated polypeptide comprising the human CT antigen peptide or a functional variant thereof, and wherein the cell is nonproliferative, or the agent is a cell expressing an isolated polypeptide comprising the human CT antigen peptide or a functional variant thereof, and wherein the cell expresses an HLA molecule that binds the polypeptide. Preferably the isolated polypeptide comprises a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In certain other embodiments, the agent is at least two, at least three, at least four or at least five different polypeptides, each coding for a different human CT antigen peptide or functional variant thereof, wherein at least one of the human CT antigen peptides is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. Preferably the at least one of the human CT antigen peptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56), or a fragment thereof.

In yet other embodiments, the agent is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:1, a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:3 or a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

Preferred cells express one or both of the polypeptide and HLA molecule recombinantly, or are nonproliferative.

In still another aspect of the invention, compositions of matter are provided that include an isolated agent that binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. In some embodiments the agent binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:1, or SEQ ID NO:3, or SEQ ID NO: 5, or SEQ ID NO:7, or SEQ ID NO:9, or SEQ ID NO:63 or the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In other embodiments, the agent is a plurality of different agents that bind selectively at least two, at least three, at least four, or at least five different such polypeptides. Preferably the at least one of the polypeptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56), or a fragment thereof.

In further embodiments, the agent is an antibody.

According to another aspect of the invention, composition of matters including a conjugate of the foregoing agents and a therapeutic or diagnostic agent are provided. Preferably the therapeutic or diagnostic is a toxin.

According to yet another aspect of the invention, pharmaceutical compositions are provided. The compositions include an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and a pharmaceutically acceptable carrier. Preferably, the isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In some embodiments, the isolated nucleic acid molecule comprises at least two isolated nucleic acid molecules coding for two different polypeptides, each polypeptide comprising a different human CT antigen., and preferably at least one of the nucleic acid molecules comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In other embodiments, the pharmaceutical compositions further include an expression vector with a promoter operably linked to the isolated nucleic acid molecule or a host cell recombinantly expressing the isolated nucleic acid molecule.

According to another aspect of the invention, pharmaceutical compositions are provided that include an isolated polypeptide comprising a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and a pharmaceutically acceptable carrier.

In certain embodiments, the isolated polypeptide comprises a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56). Preferably the isolated polypeptide comprises at least two different polypeptides, each comprising a different human CT antigen. More preferably at least one of the polypeptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56). In other preferred embodiments, the compositions include an adjuvant.

According to still another aspect of the invention, protein microarrays are provided that include at least one polypeptide encoded by a nucleic acid molecule comprising a

nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or an antigenic fragment thereof.

According to another aspect of the invention, protein microarrays are provided that include an antibody or an antigen-binding fragment thereof that specifically binds at least one polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or an antigenic fragment thereof.

According to still another aspect of the invention, nucleic acid microarrays are provided that include at least one nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample. Also provided according to the invention are, isolated fragments of a human CT antigen which, or a portion of which, binds a HLA molecule or a human antibody, wherein the CT antigen is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. In some embodiment, the fragment is part of a complex with the HLA molecule, or the fragment is between 8 and 12 amino acids in length.

According to another aspect of the invention, kits for detecting the expression of a human CT antigen are provided. The kits include a pair of isolated nucleic acid molecules each of which consists essentially of a molecule selected from the group consisting of (a) a 12-32 nucleotide contiguous segment of the nucleotide sequence of any of SEQ ID NOS:1, 3, 5, 7, 9 and 63 and (b) complements of (a), wherein the contiguous segments are nonoverlapping.

In some embodiments, the pair of isolated nucleic acid molecules is constructed and arranged to selectively amplify an isolated nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

According to yet another aspect of the invention, methods for treating a subject with a disorder characterized by expression of a human CT antigen are provided. The methods include administering to the subject an amount of an agent, which enriches selectively in the subject the presence of complexes of a HLA molecule and a human CT antigen peptide, effective to ameliorate the disorder, wherein the human CT antigen peptide is a fragment of a human CT antigen encoded by a nucleic acid molecule comprising a nucleotide sequence

selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. In some embodiments, the disorder is characterized by expression of a plurality of human CT antigens and wherein the agent is a plurality of agents, each of which enriches selectively in the subject the presence of complexes of an HLA molecule and a different human CT antigen peptide, wherein at least one of the human CT antigens is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. Preferably, at least one of the human CT antigen peptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56), or a fragment thereof. In other embodiments, the plurality is at least 2, at least 3, at least 4, or at least 5 such agents. In certain other embodiments, the agent is an isolated polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. Preferably, the disorder is cancer.

According to another aspect of the invention, methods for treating a subject having a condition characterized by expression of a human CT antigen in cells of the subject are provided. The methods include (i) removing an immunoreactive cell containing sample from the subject, (ii) contacting the immunoreactive cell containing sample to the host cell under conditions favoring production of cytolytic T cells against a human CT antigen peptide that is a fragment of the human CT antigen, (iii) introducing the cytolytic T cells to the subject in an amount effective to lyse cells which express the human CT antigen, wherein the host cell is transformed or transfected with an expression vector comprising an isolated nucleic acid molecule operably linked to a promoter, wherein the isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. Preferably the host cell recombinantly or endogenously expresses an HLA molecule which binds the human CT antigen peptide.

According to still another aspect of the invention, methods for treating a subject having a condition characterized by expression of a human CT antigen in cells of the subject are provided. The methods include (i) identifying a nucleic acid molecule expressed by the cells associated with said condition, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63; (ii) transfecting a host cell with a nucleic acid selected from the group consisting of (a) the nucleic acid molecule identified, (b) a fragment of the nucleic acid identified which includes

a segment coding for a human CT antigen, (c) deletions, substitutions or additions to (a) or (b), and (d) degenerates of (a), (b), or (c); (iii) culturing said transfected host cells to express the transfected nucleic acid molecule, and; (iv) introducing an amount of said host cells or an extract thereof to the subject effective to increase an immune response against the cells of the subject associated with the condition. Preferably the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

In some embodiments, the method also includes identifying an MHC molecule which presents a portion of an expression product of the nucleic acid molecule, wherein the host cell expresses the same MHC molecule as identified and wherein the host cell presents an MHC binding portion of the expression product of the nucleic acid molecule.

In other embodiments, the immune response comprises a B-cell response or a T cell response. Preferably, the immune response is a T-cell response which comprises generation of cytolytic T-cells specific for the host cells presenting the portion of the expression product of the nucleic acid molecule or cells of the subject expressing the human CT antigen.

In still other embodiments, the nucleic acid molecule is selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63. In certain other embodiments, the methods include treating the host cells to render them non-proliferative.

According to another aspect of the invention, methods for treating or diagnosing or monitoring a subject having a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta are provided. The methods include administering to the subject an antibody which specifically binds to the protein or a peptide derived therefrom, the antibody being coupled to a therapeutically useful agent, in an amount effective to treat the condition. Preferably the antibody is a monoclonal antibody, particularly a human monoclonal, a chimeric antibody or a humanized antibody.

According to a further aspect of the invention, methods for treating a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta are provided. The methods include administering to a subject a pharmaceutical composition of any one of claims 16-31 and 44-54 in an amount effective to prevent, delay the onset of, or inhibit the condition in the subject.

Preferably the condition is cancer. In some embodiments the methods also include first identifying that the subject expresses in a tissue abnormal amounts of the protein.

According to another aspect of the invention, methods for treating a subject having a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta are provided. The methods include (i) identifying cells from the subject which express abnormal amounts of the protein; (ii) isolating a sample of the cells; (iii) cultivating the cells, and (iv) introducing the cells to the subject in an amount effective to provoke an immune response against the cells. In some embodiments, the methods also include rendering the cells non-proliferative, prior to introducing them to the subject.

According to still another aspect of the invention, methods for treating a pathological cell condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta are provided. The methods include administering to a subject in need thereof an effective amount of an agent which inhibits the expression or activity of the protein. Preferably the agent is an inhibiting antibody which selectively binds to the protein and wherein the antibody is a monoclonal antibody, a chimeric antibody, a humanized antibody or an antibody fragment, or an antisense nucleic acid molecule which selectively binds to the nucleic acid molecule which encodes the protein. In preferred embodiments, the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1, or SEQ ID NO:3 or the nucleotide sequence of *RXF4-C* amplified by the C1 primer pair (SEQ ID NOs: 55, 56).

According to another aspect of the invention, compositions of matter useful in stimulating an immune response to a plurality of a proteins encoded by nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 are provided. The compositions include a plurality of peptides derived from the amino acid sequences of the proteins, wherein the peptides bind to one or more MHC molecules presented on the surface of cells which are not testis, fetal ovary or placenta. In some embodiments, at least a portion of the plurality of peptides bind to MHC molecules and elicit a cytolytic response thereto. In other embodiments, at least one of the proteins is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3 and the nucleotide sequence of *RXF4-C* amplified by

the C1 primer pair (SEQ ID NOs: 55, 56). Preferably the compositions further include an adjuvant, particularly a saponin, GM-CSF, or an interleukin.

In other embodiments, the compositions include at least one peptide useful in stimulating an immune response to at least one protein which is not encoded by SEQ ID NOS:1, 3, 5, 7, 9 and 63, wherein the at least one peptide binds to one or more MHC molecules.

According to another aspect of the invention, an isolated antibody is provided which selectively binds to a complex of: (i) a peptide derived from a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 and (ii) and an MHC molecule to which binds the peptide to form the complex, wherein the isolated antibody does not bind to (i) or (ii) alone. Preferably the antibody is a monoclonal antibody, a chimeric antibody, a humanized antibody, or a fragment thereof.

According to yet another aspect of the invention, methods for identifying nucleic acids that encode a CT antigen are provided. The methods include screening sequence database records for sequences that are expressed in a first set of samples consisting of cancers of at least two tissues and are expressed in a second set of samples consisting of at least one tissue selected from the group consisting of testis, ovary and placenta, and identifying as CT antigens the sequences that match the expression criteria. In preferred embodiments, the second tissue is testis only, or ovary only (preferably fetal ovary).

In other aspects of the invention, the expression criteria include cancer-specific expression and any one of: gamete-specific gene products, gene products associated with meiosis, and trophoblast-specific gene products.

In preferred embodiments of the screening methods, the sequences are expressed in cancers at least three tissues. In embodiments of the foregoing screening methods, it is preferred that the methods include a step of verification of the expression pattern of the sequences in normal tissue samples and/or tumor samples. Preferably the expression pattern is verified by nucleic acid amplification or nucleic acid hybridization.

The invention also involves the use of the genes, gene products, fragments thereof, agents which bind thereto, and so on in the preparation of medicaments. A particular medicament is for treating cancer.

These and other aspects of the invention will be described in further detail in connection with the detailed description of the invention.

## Brief Description of Figures

Fig. 1 depicts the two-step real-time RT-PCR performed to determine expression of *NY-ESO-1*, and sperm protein mRNAs in 16 normal tissues using ABI PRISM 7700 Sequence Detection System. (A) shows the real-time amplification plot. Shown is Rn (the normalized reporter signal minus the base line signal) as a function of PCR cycle number. Duplicate samples for each tissue were examined. Lines indicate each sample. The horizontal line is the threshold for detection. (B) provides the Ct (threshold cycles) values for normal tissues obtained in (A) were plotted.

Fig. 2 provides the relative mRNA expression values (n) in normal tissues standardized by the expression of β-actin. Testis specific expression was observed with *NY-ESO-1*, *SP-10*, *SP17*, *acrosin*, *PH-20*, *OY-TES-1*, *AKAP110*, ASP, ropporin, and *NYD-sp10*. Ubiquitous expression was observed with *CS-1* and *SPAG9*.

Fig. 3 is a diagram of the genomic structure of *RFX4* and alternatively spliced transcripts. Exons and introns are shown in boxes and lines, respectively. The exon/intron structure is determined according to the NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map). In alternatively spliced transcripts, the open reading frames are shown. *RFX4-A* (GenBank accession number AB044245) (SEQ ID NO:9, 10) is described by Morotomi-Yano et al. (*J. Biol. Chem.* 277(1): 836-842, 2002). *RFX4-B* (SEQ ID NO:7, 8) is also known as *NYD-sp10* (GenBank accession number AF332192). Primers used for PCR amplification are indicated by arrows.

Fig. 4 is a schematic representation of the RFX4 proteins. The DNA binding domain (DBD), the dimerization domains (DIM) and two additional conserved regions B and C are indicated.

Figs. 5A and 5B are digitized photographs of agarose gels that depict the RT-PCR analysis of *RFX4* mRNA in normal tissues (Fig. 5A) and tumors (Fig. 5B). RT-PCR was performed using the common primer pair (NYD-S and NYD-AS, shown in Fig. 3) at 30 cycle amplification. PCR products were analyzed by agarose gel electrophoresis. The same cDNA samples were tested for β-actin as an internal control.

Fig. 6 provides the expression level of *RFX4* splice variants in glioma. Primer pairs A1, A2, B1, B2, and C1 (see Fig. 3 and Table 7) were used to analyze the expression of three

alternatively spliced transcripts in gliomas and normal testis. Representative results for 3 astrocytomas G III, 3 astrocytomas G IV, and a normal testis sample are shown.

## Detailed Description of the Invention

As a consequence of T cell epitope cloning and SEREX analysis, a growing number of cancer-testis (CT) antigens have now been defined. See Table 1 and references cited therein. There are now 14 genes or gene families identified that code for presumptive cancer-testis antigens.

| CT* | System | # Genes | Chromosome Location | Detection System** | Refs. |
|---|---|---|---|---|---|
| 1 | MAGE | 16 | Xq28/Xp21 | T, Ab | 7, 10, 12, 13 |
| 2 | BAGE | 2 | Unknown | T | 14 |
| 3 | GAGE | 9 | Xp11 | T | 15, 16 |
| 4 | SSX | >5 | Xp11 | Ab | 10, 17 |
| 5 | NY-ESO-1 LAGE-1 | 2 | Xq28 | Ab, T, RDA | 18, 19 |
| 6 | SCP-1 | 3 | 1p12-p13 | Ab | 20 |
| 7 | CT7/MAGE-C1 | 1 | Xq26 | Ab, RDA | 21, 22 |
| 8 | CT8 | 1 | Unknown | Ab | 23 |
| 9 | CT9 | 1 | 1p | Ab | 24 |
| 10 | CT10/MAGE-C2 | 1 | Xq27 | RDA, Ab | 25, 26 |
| 11 | CT11p | 1 | Xq26- Xq27 | *** | 27 |
| 12 | SAGE | 1 | Xq28 | RDA | 28 |
| 13 | cTAGE-1 | 1 | 18p11 | Ab | 29 |
| 14 | OY-TES-1 | 2 | 12p12-p13 | Ab | 30 |

* Numbered according to the CT nomenclature proposed by Old & Chen (11).
** Ab=Antibody, T=CD8+ T cell, RDA=representational difference analysis.
*** Defined by differential mRNA expression in a parental vs. metastatic melanoma cell variant.

A thorough analysis of these gene reveals that they encode products with the following characteristics.

i) mRNA expression in normal tissues is restricted to testis, fetal ovary, and placenta, with little or no expression detected in adult ovary.

ii) mRNA expression in cancers of diverse origin is common - up to 30-40% of a number of different cancer types, e.g., melanoma, bladder cancer, sarcoma express one or more CT antigens.

iii) The X chromosome codes for the majority of CT antigens, but a number of more recently defined CT coding genes have a non-X chromosomal locus.

iv) In normal adult testis, expression of CT antigens is primarily restricted to immature germ cells -, e.g., spermatogonia (31). However, a recently defined CT antigen, OY-TES-1, is clearly involved in late stages of sperm maturation (see below). In fetal ovary, immature germ cells (oogonia/primary oocytes) express CT antigens, whereas oocytes in the resting primordial follicles do not (32). In fetal placenta, both cytotrophoblast and syncytiotrophoblast express CT antigens, but in term placenta, CT antigen expression is weak or absent (33).

v) A highly variable pattern of CT antigen expression is found in different cancers, from tumors showing only single positive cells or small cluster of positive cells to other tumors with a generally homogeneous expression pattern (31, 34).

vi) The function of most CT antigens is unknown, although some role in regulating gene expression appears likely. Two CT antigens, however, have known roles in gamete development - SCP-1, the synaptonemal complex protein, is involved in chromosomal reduction during meiosis (35), and OY-TES-1 is a proacrosin binding protein sp32 precursor thought to be involved in packaging acrosin in the acrosome in the sperm head (36).

vii) There is increasing evidence that CT expression is correlated with tumor progression and with tumors of higher malignant potential. For instance, a higher frequency of MAGE mRNA expression is found in metastatic vs. primary melanoma (37) and in invasive vs. superficial bladder cancer (38), and NY-ESO-1 expression in bladder cancer is correlated with high nuclear grade (39).

viii) There appears to be considerable variation in the inherent immunogenicity of different CT antigens as indicated by specific $CD8^+$ T cell and antibody responses in patients with antigen positive tumors. To date, NY-ESO-1 appears to have the strongest spontaneous immunogenicity of any of the CT antigens - e.g., up to 50% of patients with advanced NY-ESO-$1^+$ tumors develop humoral and cellular immunity to NY-ESO-1 (40, 41).

These characteristics indicate the desirability of cancer-testis antigens for use in diagnostics and therapeutics. These characteristics also provide a basis for the identification of additional cancer-testis antigens.

While others have attempted to identify cancer related sequences in public databases by the use of bioinformatics techniques, (e.g., database mining plus rapid screening by fluorescent-PCR expression, Loging et al., *Genome Res* 10(9):1393-402, 2000), these techniques have not focused on the identification of nucleic acid sequences that fit the preferred cancer-testis antigen profile. In particular, the present invention includes the

identification of cancer-testis sequences by more stringent criteria. The database analysis criteria for identifying cancer-testis antigen sequences include the requirement that the sequences are expressed in cancers from at least two different tissues, and preferably are expressed in cancers from at least three different tissues. In addition, the sequences preferably have normal tissue expression restricted to one or more tissue selected from the group consisting of testis, placenta and ovary (preferably only fetal ovary).

In the above summary and in the ensuing description, lists of sequences are provided. The lists are meant to embrace each single sequence separately, two or more sequences together where they form a part of the same gene, any combination of two or more sequences which relate to different genes, including and up to the total number on the list, as if each and every combination were separately and specifically enumerated. Likewise, when mentioning fragment size, it is intended that a range embrace the smallest fragment mentioned to the full-length of the sequence (less one nucleotide or amino acid so that it is a fragment), each and every fragment length intended as if specifically enumerated. Thus, if a fragment could be between 10 and 15 in length, it is explicitly meant to mean 10, 11, 12, 13, 14, or 15 in length.

The summary and the claims mention antigen precursors and antigens. As used in the summary and in the claims, a precursor is substantially the full-length protein encoded by the coding region of the isolated nucleic acid and the antigen is a peptide which complexes with MHC, preferably HLA, and which participates in the immune response as part of that complex. Such antigens are typically 9 amino acids long, although this may vary slightly.

As used herein, a subject is a human, non-human primate, cow, horse, pig, sheep, goat, dog, cat or rodent. In all embodiments human cancer antigens and human subjects are preferred.

The present invention in one aspect involves the identification of human CT antigens using autologous antisera of subjects having cancer. The sequences representing CT antigen genes identified according to the methods described herein are presented in the attached Sequence Listing. The nature of the sequences as encoding CT antigens recognized by the immune systems of cancer patients is, of course, unexpected.

The invention thus involves in one aspect CT antigen polypeptides, genes encoding those polypeptides, functional modifications and variants of the foregoing, useful fragments of the foregoing, as well as diagnostics and therapeutics relating thereto.

Homologs and alleles of the CT antigen nucleic acids of the invention can be identified by conventional techniques. Thus, an aspect of the invention is those nucleic acid sequences which code for CT antigen precursors.

The term "stringent conditions" as used herein refers to parameters with which the art is familiar. Nucleic acid hybridization parameters may be found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. More specifically, stringent conditions, as used herein, refers, for example, to hybridization at 65°C in hybridization buffer (3.5 x SSC, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, 0.02% Bovine Serum Albumin, 2.5mM $NaH_2PO_4$(pH7), 0.5% SDS, 2mM EDTA). SSC is 0.15M sodium chloride/0.15M sodium citrate, pH7; SDS is sodium dodecyl sulphate; and EDTA is ethylenediaminetetracetic acid. After hybridization, the membrane upon which the DNA is transferred is washed, for example, in 2 x SSC at room temperature and then at 0.1 - 0.5 x SSC/0.1 x SDS at temperatures up to 68°C.

There are other conditions, reagents, and so forth which can be used, which result in a similar degree of stringency. The skilled artisan will be familiar with such conditions, and thus they are not given here. It will be understood, however, that the skilled artisan will be able to manipulate the conditions in a manner to permit the clear identification of homologs and alleles of CT antigen nucleic acids of the invention (e.g., by using lower stringency conditions). The skilled artisan also is familiar with the methodology for screening cells and libraries for expression of such molecules which then are routinely isolated, followed by isolation of the pertinent nucleic acid molecule and sequencing.

In general homologs and alleles typically will share at least 75% nucleotide identity and/or at least 90% amino acid identity to the sequences of CT antigen nucleic acid and polypeptides, respectively, in some instances will share at least 90% nucleotide identity and/or at least 95% amino acid identity and in still other instances will share at least 95% nucleotide identity and/or at least 99% amino acid identity. The homology can be calculated using various, publicly available software tools developed by NCBI (Bethesda, Maryland) that can be obtained through the internet (ftp:/ncbi.nlm.nih.gov/pub/). Exemplary tools include the BLAST software available at http://www.ncbi.nlm.nih.gov, using default settings. Pairwise and ClustalW alignments (BLOSUM30 matrix setting) as well as Kyte-Doolittle hydropathic analysis can be obtained using the MacVector sequence analysis software

(Oxford Molecular Group). Watson-Crick complements of the foregoing nucleic acids also are embraced by the invention.

In screening for CT antigen genes, a Southern blot may be performed using the foregoing conditions, together with a radioactive probe. After washing the membrane to which the DNA is finally transferred, the membrane can be placed against X-ray film to detect the radioactive signal. In screening for the expression of CT antigen nucleic acids, Northern blot hybridizations using the foregoing can be performed on samples taken from cancer patients or subjects suspected of having a condition characterized by expression of CT antigen genes. Amplification protocols such as polymerase chain reaction using primers which hybridize to the sequences presented also can be used for detection of the CT antigen genes or expression thereof.

The invention also includes degenerate nucleic acids which include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Thus, it will be apparent to one of ordinary skill in the art that any of the serine-encoding nucleotide triplets may be employed to direct the protein synthesis apparatus, *in vitro* or *in vivo*, to incorporate a serine residue into an elongating CT antigen polypeptide. Similarly, nucleotide sequence triplets which encode other amino acid residues include, but are not limited to: CCA, CCC, CCG and CCT (proline codons); CGA, CGC, CGG, CGT, AGA and AGG (arginine codons); ACA, ACC, ACG and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon sequence due to the degeneracy of the genetic code.

The invention also provides modified nucleic acid molecules which include additions, substitutions and deletions of one or more nucleotides. In preferred embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as antigenicity, enzymatic activity, receptor binding, formation of complexes by binding of peptides by MHC class I and class II molecules, etc. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, preferably polypeptides having conservative amino acid substitutions as are described elsewhere herein. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules and

in preferred embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

For example, modified nucleic acid molecules which encode polypeptides having single amino acid changes can be prepared. Each of these nucleic acid molecules can have one, two or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules which encode polypeptides having two amino acid changes can be prepared which have, e.g., 2-6 nucleotide changes. Numerous modified nucleic acid molecules like these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions which code for the amino acid substitutions. Additional nucleic acid molecules that encode polypeptides having additional substitutions (i.e., 3 or more), additions or deletions (e.g., by introduction of a stop codon or a splice site(s)) also can be prepared and are embraced by the invention as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acids or polypeptides can be tested by routine experimentation for retention of structural relation or activity to the nucleic acids and/or polypeptides disclosed herein.

The invention also provides isolated fragments of CT antigen nucleic acid sequences or complements thereof, and in particular unique fragments. A unique fragment is one that is a 'signature' for the larger nucleic acid. It, for example, is long enough to assure that its precise sequence is not found in molecules within the human genome outside of the CT antigen nucleic acids defined above (and human alleles). Those of ordinary skill in the art may apply routine procedures to determine if a fragment is unique within the human genome, such as the use of publicly available sequence comparison software to selectively distinguish the sequence fragment of interest from other sequences in the human genome, although *in vitro* confirmatory hybridization and sequencing analysis may be performed.

Fragments can be used as probes in Southern and Northern blot assays to identify CT antigen nucleic acids, or can be used in amplification assays such as those employing PCR. As known to those skilled in the art, large probes such as 200, 250, 300 or more nucleotides are preferred for certain uses such as Southern and Northern blots, while smaller fragments will be preferred for uses such as PCR. Fragments also can be used to produce fusion

proteins for generating antibodies or determining binding of the polypeptide fragments, or for generating immunoassay components. Likewise, fragments can be employed to produce nonfused fragments of the CT antigen polypeptides, useful, for example, in the preparation of antibodies, and in immunoassays. Fragments further can be used as antisense molecules to inhibit the expression of CT antigen nucleic acids and polypeptides, particularly for therapeutic purposes as described in greater detail below.

As mentioned above, this disclosure intends to embrace each and every fragment of each sequence, beginning at the first nucleotide, the second nucleotide and so on, up to 8 nucleotides short of the end, and ending anywhere from nucleotide number 8, 9, 10 and so on for each sequence, up to the entire length of the disclosed sequence. Preferred fragments are those useful as amplification primers, e.g., typically between 12 and 32 nucleotides (e.g. 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 and 32) in length.

Those skilled in the art are well versed in methods for selecting such sequences, typically on the basis of the ability of the fragment to selectively distinguish the sequence of interest from other sequences in the human genome of the fragment to those on known databases typically is all that is necessary, although *in vitro* confirmatory hybridization and sequencing analysis may be performed.

Especially preferred fragments include nucleic acids encoding a series of epitopes, known as "polytopes". The epitopes can be arranged in sequential or overlapping fashion (*see, e.g.*, Thomson et al., *Proc. Natl. Acad. Sci. USA* 92:5845-5849, 1995; Gilbert et al., *Nature Biotechnol.* 15:1280-1284, 1997), with or without the natural flanking sequences, and can be separated by unrelated linker sequences if desired. The polytope is processed to generated individual epitopes which are recognized by the immune system for generation of immune responses.

Thus, for example, peptides derived from a polypeptide having an amino acid sequence encoded by one of the nucleic acid disclosed herein, and which are presented by MHC molecules and recognized by CTL or T helper lymphocytes, can be combined with peptides from one or more other CT antigens (e.g. by preparation of hybrid nucleic acids or polypeptides) to form "polytopes". The two or more peptides (or nucleic acids encoding the peptides) can be selected from those described herein, or they can include one or more peptides of previously known CT antigens. Exemplary cancer associated peptide antigens that can be administered to induce or enhance an immune response are derived from tumor associated genes and encoded proteins including MAGE-A1, MAGE-A2, MAGE-A3,

MAGE-A4, MAGE-A5, MAGE-A6, MAGE-A7, MAGE-A8, MAGE-A9, MAGE-A10, MAGE-A11, MAGE-A12, GAGE-1, GAGE-2, GAGE-3, GAGE-4, GAGE-5, GAGE-6, GAGE-7, GAGE-8, GAGE-9, BAGE-1, RAGE-1, LB33/MUM-1, PRAME, NAG, MAGE-B2, MAGE-B3, MAGE-B4, tyrosinase, brain glycogen phosphorylase, Melan-A, MAGE-C1, MAGE-C2, MAGE-C3, MAGE-C4, MAGE-C5, NY-ESO-1, LAGE-1, SSX-1, SSX-2 (HOM-MEL-40), SSX-4, SSX-5, SCP-1 and CT-7. See, for example, PCT application publication no. WO96/10577. Other examples will be known to one of ordinary skill in the art and can be used in the invention in a like manner as those disclosed herein. Other examples of HLA class I and HLA class II binding peptides will be known to one of ordinary skill in the art. For example, see the following references: Coulie, *Stem Cells* 13:393-403, 1995; Traversari et al., *J. Exp. Med.* 176:1453-1457, 1992; Chaux et al., *J. Immunol.* 163:2928-2936, 1999; Fujie et al., *Int. J. Cancer* 80:169-172, 1999; Tanzarella et al., *Cancer Res.* 59:2668-2674, 1999; van der Bruggen et al., *Eur. J. Immunol.* 24:2134-2140, 1994; Chaux et al., *J. Exp. Med.* 189:767-778, 1999; Kawashima et al, *Hum. Immunol.* 59:1-14, 1998; Tahara et al., *Clin. Cancer Res.* 5:2236-2241, 1999; Gaugler et al., *J. Exp. Med.* 179:921-930, 1994; van der Bruggen et al., *Eur. J. Immunol.* 24:3038-3043, 1994; Tanaka et al., *Cancer Res.* 57:4465-4468, 1997; Oiso et al., *Int. J. Cancer* 81:387-394, 1999; Herman et al., *Immunogenetics* 43:377-383, 1996; Manici et al., *J. Exp. Med.* 189:871-876, 1999; Duffour et al., *Eur. J. Immunol.* 29:3329-3337, 1999; Zorn et al., *Eur. J. Immunol.* 29:602-607, 1999; Huang et al., *J. Immunol.*162:6849-6854, 1999; Boël et al., *Immunity* 2:167-175, 1995; Van den Eynde et al., *J. Exp. Med.* 182:689-698, 1995; De Backer et al., *Cancer Res.* 59:3157-3165, 1999; Jäger et al., *J. Exp. Med.* 187:265-270, 1998; Wang et al., *J. Immunol.* 161:3596-3606, 1998; Aarnoudse et al., *Int. J. Cancer* 82:442-448, 1999; Guilloux et al., *J. Exp. Med.* 183:1173-1183, 1996; Lupetti et al., *J. Exp. Med.* 188:1005-1016, 1998; Wölfel et al., *Eur. J. Immunol.* 24:759-764, 1994; Skipper et al., *J. Exp. Med.* 183:527-534, 1996; Kang et al., *J. Immunol.* 155:1343-1348, 1995; Morel et al., *Int. J. Cancer* 83:755-759, 1999; Brichard et al., *Eur. J. Immunol.* 26:224-230, 1996; Kittlesen et al., *J. Immunol.* 160:2099-2106, 1998; Kawakami et al., *J. Immunol.* 161:6985-6992, 1998; Topalian et al., *J. Exp. Med.* 183:1965-1971, 1996; Kobayashi et al., *Cancer Research* 58:296-301, 1998; Kawakami et al., *J. Immunol.* 154:3961-3968, 1995; Tsai et al., *J. Immunol.* 158:1796-1802, 1997; Cox et al., *Science* 264:716-719, 1994; Kawakami et al., *Proc. Natl. Acad. Sci. USA* 91:6458-6462, 1994; Skipper et al., *J. Immunol.* 157:5027-5033, 1996; Robbins et al., *J. Immunol.* 159:303-308, 1997; Castelli et al, *J. Immunol.* 162:1739-1748, 1999; Kawakami et al., *J. Exp. Med.*

180:347-352, 1994; Castelli et al., *J. Exp. Med.* 181:363-368, 1995; Schneider et al., *Int. J. Cancer* 75:451-458, 1998; Wang et al., *J. Exp. Med.* 183:1131-1140, 1996; Wang et al., *J. Exp. Med.* 184:2207-2216, 1996; Parkhurst et al., *Cancer Research* 58:4895-4901, 1998; Tsang et al., *J. Natl Cancer Inst* 87:982-990, 1995; Correale et al., *J Natl Cancer Inst* 89:293-300, 1997; Coulie et al., *Proc. Natl. Acad. Sci. USA* 92:7976-7980, 1995; Wölfel et al., *Science* 269:1281-1284, 1995; Robbins et al., *J. Exp. Med.* 183:1185-1192, 1996; Brändle et al., *J. Exp. Med.* 183:2501-2508, 1996; ten Bosch et al., *Blood* 88:3522-3527, 1996; Mandruzzato et al., *J. Exp. Med.* 186:785-793, 1997; Guéguen et al., *J. Immunol.* 160:6188-6194, 1998; Gjertsen et al., *Int. J. Cancer* 72:784-790, 1997; Gaudin et al., *J. Immunol.* 162:1730-1738, 1999; Chiari et al., *Cancer Res.* 59:5785-5792, 1999; Hogan et al., *Cancer Res.* 58:5144-5150, 1998; Pieper et al., *J. Exp. Med.* 189:757-765, 1999; Wang et al., *Science* 284:1351-1354, 1999; Fisk et al., *J. Exp. Med.* 181:2109-2117, 1995; Brossart et al., *Cancer Res.* 58:732-736, 1998; Röpke et al., *Proc. Natl. Acad. Sci. USA* 93:14704-14707, 1996; Ikeda et al., *Immunity* 6:199-208, 1997; Ronsin et al., *J. Immunol.* 163:483-490, 1999; Vonderheide et al., *Immunity* 10:673-679,1999.

One of ordinary skill in the art can prepare polypeptides comprising one or more CT antigen peptides and one or more of the foregoing cancer associated peptides, or nucleic acids encoding such polypeptides, according to standard procedures of molecular biology.

Thus polytopes are groups of two or more potentially immunogenic or immune response stimulating peptides which can be joined together in various arrangements (e.g. concatenated, overlapping). The polytope (or nucleic acid encoding the polytope) can be administered in a standard immunization protocol, e.g. to animals, to test the effectiveness of the polytope in stimulating, enhancing and/or provoking an immune response.

The peptides can be joined together directly or via the use of flanking sequences to form polytopes, and the use of polytopes as vaccines is well known in the art (see, e.g., Thomson et al., *Proc. Acad. Natl. Acad. Sci USA* 92(13):5845-5849, 1995; Gilbert et al., *Nature Biotechnol.* 15(12):1280-1284, 1997; Thomson et al., *J. Immunol.* 157(2):822-826, 1996; Tam et al., *J. Exp. Med.* 171(1):299-306, 1990). For example, Tam showed that polytopes consisting of both MHC class I and class II binding epitopes successfully generated antibody and protective immunity in a mouse model. Tam also demonstrated that polytopes comprising "strings" of epitopes are processed to yield individual epitopes which are presented by MHC molecules and recognized by CTLs. Thus polytopes containing various

numbers and combinations of epitopes can be prepared and tested for recognition by CTLs and for efficacy in increasing an immune response.

It is known that tumors express a set of tumor antigens, of which only certain subsets may be expressed in the tumor of any given patient. Polytopes can be prepared which correspond to the different combination of epitopes representing the subset of tumor rejection antigens expressed in a particular patient. Polytopes also can be prepared to reflect a broader spectrum of tumor rejection antigens known to be expressed by a tumor type. Polytopes can be introduced to a patient in need of such treatment as polypeptide structures, or via the use of nucleic acid delivery systems known in the art (see, e.g., Allsopp et al., *Eur. J. Immunol.* 26(8):1951-1959, 1996). Adenovirus, pox viruses, Ty-virus like particles, adeno-associated virus, alphaviruses, plasmids, bacteria, etc. can be used in such delivery. One can test the polytope delivery systems in mouse models to determine efficacy of the delivery system. The systems also can be tested in human clinical trials.

In instances in which a human HLA class I molecule presents tumor rejection antigens derived from CT antigens, the expression vector may also include a nucleic acid sequence coding for the HLA molecule that presents any particular tumor rejection antigen derived from these nucleic acids and polypeptides. Alternatively, the nucleic acid sequence coding for such a HLA molecule can be contained within a separate expression vector. In a situation where the vector contains both coding sequences, the single vector can be used to transfect a cell which does not normally express either one. Where the coding sequences for a CT antigen precursor and the HLA molecule which presents it are contained on separate expression vectors, the expression vectors can be cotransfected. The CT antigen precursor coding sequence may be used alone, when, e.g. the host cell already expresses a HLA molecule which presents a CT antigen derived from precursor molecules. Of course, there is no limit on the particular host cell which can be used. As the vectors which contain the two coding sequences may be used in any antigen-presenting cells if desired, and the gene for CT antigen precursor can be used in host cells which do not express a HLA molecule which presents a CT antigen. Further, cell-free transcription systems may be used in lieu of cells.

As mentioned above, the invention embraces antisense oligonucleotides that selectively bind to a nucleic acid molecule encoding a CT antigen polypeptide, to reduce the expression of CT antigens. This is desirable in virtually any medical condition wherein a reduction of expression of CT antigens is desirable, e.g., in the treatment of cancer. This is

also useful for *in vitro* or *in vivo* testing of the effects of a reduction of expression of one or more CT antigens.

As used herein, the term "antisense oligonucleotide" or "antisense" describes an oligonucleotide that is an oligoribonucleotide, oligodeoxyribonucleotide, modified oligoribonucleotide, or modified oligodeoxyribonucleotide which hybridizes under physiological conditions to DNA comprising a particular gene or to an mRNA transcript of that gene and, thereby, inhibits the transcription of that gene and/or the translation of that mRNA. The antisense molecules are designed so as to interfere with transcription or translation of a target gene upon hybridization with the target gene or transcript. Those skilled in the art will recognize that the exact length of the antisense oligonucleotide and its degree of complementarity with its target will depend upon the specific target selected, including the sequence of the target and the particular bases which comprise that sequence. It is preferred that the antisense oligonucleotide be constructed and arranged so as to bind selectively with the target under physiological conditions, i.e., to hybridize substantially more to the target sequence than to any other sequence in the target cell under physiological conditions. Based upon the sequences of nucleic acids encoding CT antigens, or upon allelic or homologous genomic and/or cDNA sequences, one of skill in the art can easily choose and synthesize any of a number of appropriate antisense molecules for use in accordance with the present invention.

In order to be sufficiently selective and potent for inhibition, such antisense oligonucleotides should comprise at least 10 and, more preferably, at least 15 consecutive bases which are complementary to the target, although in certain cases modified oligonucleotides as short as 7 bases in length have been used successfully as antisense oligonucleotides (Wagner et al., *Nature Biotechnol.* 14:840-844, 1996). Most preferably, the antisense oligonucleotides comprise a complementary sequence of 20-30 bases.

Although oligonucleotides may be chosen which are antisense to any region of the gene or mRNA transcripts, in preferred embodiments the antisense oligonucleotides correspond to N-terminal or 5' upstream sites such as translation initiation, transcription initiation or promoter sites. In addition, 3'-untranslated regions may be targeted. Targeting to mRNA splicing sites has also been used in the art but may be less preferred if alternative mRNA splicing occurs. In addition, the antisense is targeted, preferably, to sites in which mRNA secondary structure is not expected (see, e.g., Sainio et al., *Cell Mol. Neurobiol.* 14(5):439-457, 1994) and at which proteins are not expected to bind. Suitable antisense

molecules can be identified by a "gene walk" experiment in which overlapping oligonucleotides corresponding to the CT antigen nucleic acid are synthesized and tested for the ability to inhibit expression, cause the degradation of sense transcripts, etc. Finally, although the listed sequences are cDNA sequences, one of ordinary skill in the art may easily derive the genomic DNA corresponding to the cDNA of a CT antigen. Thus, the present invention also provides for antisense oligonucleotides which are complementary to the genomic DNA corresponding to nucleic acids encoding CT antigens. Similarly, antisense to allelic or homologous cDNAs and genomic DNAs are enabled without undue experimentation.

In one set of embodiments, the antisense oligonucleotides of the invention may be composed of "natural" deoxyribonucleotides, ribonucleotides, or any combination thereof. That is, the 5' end of one native nucleotide and the 3' end of another native nucleotide may be covalently linked, as in natural systems, via a phosphodiester internucleoside linkage. These oligonucleotides may be prepared by art recognized methods which may be carried out manually or by an automated synthesizer. They also may be produced recombinantly by vectors.

In preferred embodiments, however, the antisense oligonucleotides of the invention also may include "modified" oligonucleotides. That is, the oligonucleotides may be modified in a number of ways which do not prevent them from hybridizing to their target but which enhance their stability or targeting or which otherwise enhance their therapeutic effectiveness.

The term "modified oligonucleotide" as used herein describes an oligonucleotide in which (1) at least two of its nucleotides are covalently linked via a synthetic internucleoside linkage (i.e., a linkage other than a phosphodiester linkage between the 5' end of one nucleotide and the 3' end of another nucleotide) and/or (2) a chemical group not normally associated with nucleic acids has been covalently attached to the oligonucleotide. Preferred synthetic internucleoside linkages are phosphorothioates, alkylphosphonates, phosphorodithioates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters and peptides.

The term "modified oligonucleotide" also encompasses oligonucleotides with a covalently modified base and/or sugar. For example, modified oligonucleotides include oligonucleotides having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a

phosphate group at the 5' position. Thus modified oligonucleotides may include a 2'-O-alkylated ribose group. In addition, modified oligonucleotides may include sugars such as arabinose instead of ribose. Base analogs such as C-5 propyne modified bases also can be included (*Nature Biotechnol.* 14:840-844, 1996). The present invention, thus, contemplates pharmaceutical preparations containing modified antisense molecules that are complementary to and hybridizable with, under physiological conditions, nucleic acids encoding the CT antigen polypeptides, together with pharmaceutically acceptable carriers.

Antisense oligonucleotides may be administered as part of a pharmaceutical composition. Such a pharmaceutical composition may include the antisense oligonucleotides in combination with any standard physiologically and/or pharmaceutically acceptable carriers which are known in the art. The compositions should be sterile and contain a therapeutically effective amount of the antisense oligonucleotides in a unit of weight or volume suitable for administration to a patient. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. The term "physiologically acceptable" refers to a non-toxic material that is compatible with a biological system such as a cell, cell culture, tissue, or organism. The characteristics of the carrier will depend on the route of administration. Physiologically and pharmaceutically acceptable carriers include diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials which are well known in the art, as further described below.

As used herein, a "vector" may be any of a number of nucleic acids into which a desired sequence may be inserted by restriction and ligation for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA although RNA vectors are also available. Vectors include, but are not limited to, plasmids, phagemids and virus genomes. A cloning vector is one which is able to replicate autonomously or integrated in the genome in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in the host cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host bacterium or just a single time per host before the host reproduces by mitosis. In the case of phage, replication may occur actively during a lytic phase or passively during a lysogenic phase. An expression vector is one into which a desired DNA sequence may be inserted by restriction and ligation such that it is operably joined to regulatory

sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β-galactosidase, luciferase or alkaline phosphatase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (e.g., green fluorescent protein). Preferred vectors are those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

As used herein, a coding sequence and regulatory sequences are said to be "operably" joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of that DNA sequence such that the resulting transcript might be translated into the desired protein or polypeptide.

The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. Especially, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. The vectors of the invention may optionally include 5' leader or signal sequences. The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA (RNA) encoding a CT antigen polypeptide or fragment or variant thereof. That heterologous DNA (RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell.

Preferred systems for mRNA expression in mammalian cells are those such as pRc/CMV or pcDNA3.1 (available from Invitrogen, Carlsbad, CA) that contain a selectable marker such as a gene that confers G418 resistance (which facilitates the selection of stably transfected cell lines) and the human cytomegalovirus (CMV) enhancer-promoter sequences. Additionally, suitable for expression in primate or canine cell lines is the pCEP4 vector (Invitrogen), which contains an Epstein Barr Virus (EBV) origin of replication, facilitating the maintenance of plasmid as a multicopy extrachromosomal element. Another expression vector is the pEF-BOS plasmid containing the promoter of polypeptide Elongation Factor 1α, which stimulates efficiently transcription *in vitro*. The plasmid is described by Mishizuma and Nagata (*Nuc. Acids Res.* 18:5322, 1990), and its use in transfection experiments is disclosed by, for example, Demoulin (*Mol. Cell. Biol.* 16:4710-4716, 1996). Still another preferred expression vector is an adenovirus, described by Stratford-Perricaudet, which is defective for E1 and E3 proteins (*J. Clin. Invest.* 90:626-630, 1992). The use of the adenovirus as an Adeno.P1A recombinant for the expression of an antigen is disclosed by Warnier et al., in intradermal injection in mice for immunization against P1A (*Int. J. Cancer*, 67:303-310, 1996).

The invention also embraces so-called expression kits, which allow the artisan to prepare a desired expression vector or vectors. Such expression kits include at least separate portions of a vector and one or more of the previously discussed CT antigen nucleic acid molecules. Other components may be added, as desired, as long as the previously mentioned nucleic acid molecules, which are required, are included. The invention also includes kits for amplification of a CT antigen nucleic acid, including at least one pair of amplification primers which hybridize to a CT antigen nucleic acid. The primers preferably are 12-32 nucleotides in length and are non-overlapping to prevent formation of "primer-dimers". One of the primers will hybridize to one strand of the CT antigen nucleic acid and the second primer will hybridize to the complementary strand of the CT antigen nucleic acid, in an

arrangement which permits amplification of the CT antigen nucleic acid. Selection of appropriate primer pairs is standard in the art. For example, the selection can be made with assistance of a computer program designed for such a purpose, optionally followed by testing the primers for amplification specificity and efficiency.

The invention also permits the construction of CT antigen gene "knock-outs" and "knock-ins" in cells and in animals, providing materials for studying certain aspects of cancer and immune system responses to cancer.

The invention also provides isolated polypeptides (including whole proteins and partial proteins) encoded by the foregoing CT antigen nucleic acids. Such polypeptides are useful, for example, alone or as fusion proteins to generate antibodies, as components of an immunoassay or diagnostic assay or as therapeutics. CT antigen polypeptides can be isolated from biological samples including tissue or cell homogenates, and can also be expressed recombinantly in a variety of prokaryotic and eukaryotic expression systems by constructing an expression vector appropriate to the expression system, introducing the expression vector into the expression system, and isolating the recombinantly expressed protein. Short polypeptides, including antigenic peptides (such as are presented by MHC molecules on the surface of a cell for immune recognition) also can be synthesized chemically using well-established methods of peptide synthesis.

A unique fragment of a CT antigen polypeptide, in general, has the features and characteristics of unique fragments as discussed above in connection with nucleic acids. As will be recognized by those skilled in the art, the size of the unique fragment will depend upon factors such as whether the fragment constitutes a portion of a conserved protein domain. Thus, some regions of CT antigens will require longer segments to be unique while others will require only short segments, typically between 5 and 12 amino acids (e.g. 5, 6, 7, 8, 9, 10, 11 or 12 or more amino acids including each integer up to the full length).

Fragments of a CT antigen polypeptide preferably are those fragments which retain a distinct functional capability of the polypeptide. Functional capabilities which can be retained in a fragment of a polypeptide include interaction with antibodies, interaction with other polypeptides or fragments thereof, selective binding of nucleic acids or proteins, and enzymatic activity. One important activity is the ability to act as a signature for identifying the polypeptide. Another is the ability to complex with HLA and to provoke in a human an immune response. Those skilled in the art are well versed in methods for selecting unique amino acid sequences, typically on the basis of the ability of the fragment to selectively

distinguish the sequence of interest from non-family members. A comparison of the sequence of the fragment to those on known databases typically is all that is necessary.

The invention embraces variants of the CT antigen polypeptides described above. As used herein, a "variant" of a CT antigen polypeptide is a polypeptide which contains one or more modifications to the primary amino acid sequence of a CT antigen polypeptide. Modifications which create a CT antigen variant can be made to a CT antigen polypeptide 1) to reduce or eliminate an activity of a CT antigen polypeptide; 2) to enhance a property of a CT antigen polypeptide, such as protein stability in an expression system or the stability of protein-protein binding; 3) to provide a novel activity or property to a CT antigen polypeptide, such as addition of an antigenic epitope or addition of a detectable moiety; or 4) to provide equivalent or better binding to an HLA molecule. Modifications to a CT antigen polypeptide are typically made to the nucleic acid which encodes the CT antigen polypeptide, and can include deletions, point mutations, truncations, amino acid substitutions and additions of amino acids or non-amino acid moieties. Alternatively, modifications can be made directly to the polypeptide, such as by cleavage, addition of a linker molecule, addition of a detectable moiety, such as biotin, addition of a fatty acid, and the like. Modifications also embrace fusion proteins comprising all or part of the CT antigen amino acid sequence. One of skill in the art will be familiar with methods for predicting the effect on protein conformation of a change in protein sequence, and can thus "design" a variant CT antigen polypeptide according to known methods. One example of such a method is described by Dahiyat and Mayo in *Science* 278:82-87, 1997, whereby proteins can be designed *de novo*. The method can be applied to a known protein to vary a only a portion of the polypeptide sequence. By applying the computational methods of Dahiyat and Mayo, specific variants of a CT antigen polypeptide can be proposed and tested to determine whether the variant retains a desired conformation.

In general, variants include CT antigen polypeptides which are modified specifically to alter a feature of the polypeptide unrelated to its desired physiological activity. For example, cysteine residues can be substituted or deleted to prevent unwanted disulfide linkages. Similarly, certain amino acids can be changed to enhance expression of a CT antigen polypeptide by eliminating proteolysis by proteases in an expression system (e.g., dibasic amino acid residues in yeast expression systems in which KEX2 protease activity is present).

Mutations of a nucleic acid which encode a CT antigen polypeptide preferably preserve the amino acid reading frame of the coding sequence, and preferably do not create regions in the nucleic acid which are likely to hybridize to form secondary structures, such a hairpins or loops, which can be deleterious to expression of the variant polypeptide.

Mutations can be made by selecting an amino acid substitution, or by random mutagenesis of a selected site in a nucleic acid which encodes the polypeptide. Variant polypeptides are then expressed and tested for one or more activities to determine which mutation provides a variant polypeptide with the desired properties. Further mutations can be made to variants (or to non-variant CT antigen polypeptides) which are silent as to the amino acid sequence of the polypeptide, but which provide preferred codons for translation in a particular host. The preferred codons for translation of a nucleic acid in, e.g., *E. coli*, are well known to those of ordinary skill in the art. Still other mutations can be made to the noncoding sequences of a CT antigen gene or cDNA clone to enhance expression of the polypeptide. The activity of variants of CT antigen polypeptides can be tested by cloning the gene encoding the variant CT antigen polypeptide into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the variant CT antigen polypeptide, and testing for a functional capability of the CT antigen polypeptides as disclosed herein. For example, the variant CT antigen polypeptide can be tested for binding to antibodies or T cells. Preferred variants are those that compete for binding with the original polypeptide for binding to antibodies or T cells. Preparation of other variant polypeptides may favor testing of other activities, as will be known to one of ordinary skill in the art.

The skilled artisan will also realize that conservative amino acid substitutions may be made in CT antigen polypeptides to provide functionally equivalent variants of the foregoing polypeptides, i.e., the variants retain the functional capabilities of the CT antigen polypeptides. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Variants can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. Exemplary functionally equivalent variants of the

CT antigen polypeptides include conservative amino acid substitutions in the amino acid sequences of proteins disclosed herein. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D.

For example, upon determining that a peptide derived from a CT antigen polypeptide is presented by an MHC molecule and recognized by CTLs (e.g., as described in the Examples), one can make conservative amino acid substitutions to the amino acid sequence of the peptide, particularly at residues which are thought not to be direct contact points with the MHC molecule, i.e., the anchor residues that confer MHC binding. One of ordinary skill in the art will know these residues and will preferentially substitute other amino acid residues in the peptides in making variants. It is possible also to use other members of the consensus amino acids for a particular anchor residue. For example, consensus anchor residues for HLA-B35 are P in position 2 and Y, F, M, L or I in position 9. Therefore, if position 9 of a peptide was tyrosine (Y), one could substitute phenylalanine (F), methionine (M), leucine (L) or isoleucine (I) and maintain a consensus amino acid at the anchor residue positions of the peptide.

In general, it is preferred that fewer than all of the amino acids are changed when preparing variant polypeptides. Where particular amino acid residues are known to confer function, such amino acids will not be replaced, or alternatively, will be replaced by conservative amino acid substitutions. Preferably, 1, 2, 3, 4, 5, 6, 7, 8, and so on up to one fewer than the length of the peptide are changed when preparing variant polypeptides. It is generally preferred that the fewest number of substitutions is made. Thus, one method for generating variant polypeptides is to substitute all other amino acids for a particular single amino acid, then assay activity of the variant, then repeat the process with one or more of the polypeptides having the best activity.

As another example, methods for identifying functional variants of HLA class II binding peptides are provided in a published PCT application of Strominger and Wucherpfennig (PCT/US96/03182). Peptides bearing one or more amino acid substitutions also can be tested for concordance with known HLA/MHC motifs prior to synthesis using, e.g. the computer program described by D'Amaro and Drijfhout (D'Amaro et al., *Human Immunol.* 43:13-18, 1995; Drijfhout et al., *Human Immunol.* 43:1-12, 1995). The substituted peptides can then be tested for binding to the MHC molecule and recognition by CTLs when

bound to MHC. These variants can be tested for improved stability and are useful, *inter alia*, in vaccine compositions.

Conservative amino-acid substitutions in the amino acid sequence of CT antigen polypeptides to produce functionally equivalent variants of CT antigen polypeptides typically are made by alteration of a nucleic acid encoding a CT antigen polypeptide. Such substitutions can be made by a variety of methods known to one of ordinary skill in the art. For example, amino acid substitutions may be made by PCR-directed mutation, site-directed mutagenesis according to the method of Kunkel (Kunkel, *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985), or by chemical synthesis of a gene encoding a CT antigen polypeptide. Where amino acid substitutions are made to a small unique fragment of a CT antigen polypeptide, such as an antigenic epitope recognized by autologous or allogeneic sera or cytolytic T lymphocytes, the substitutions can be made by directly synthesizing the peptide. The activity of functionally equivalent fragments of CT antigen polypeptides can be tested by cloning the gene encoding the altered CT antigen polypeptide into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the altered CT antigen polypeptide, and testing for a functional capability of the CT antigen polypeptides as disclosed herein. Peptides which are chemically synthesized can be tested directly for function, e.g., for binding to antisera recognizing associated antigens.

The invention also provides, in certain embodiments, "dominant negative" polypeptides derived from CT antigen polypeptides. A dominant negative polypeptide is an inactive variant of a protein, which, by interacting with the cellular machinery, displaces an active protein from its interaction with the cellular machinery or competes with the active protein, thereby reducing the effect of the active protein. For example, a dominant negative receptor which binds a ligand but does not transmit a signal in response to binding of the ligand can reduce the biological effect of expression of the ligand. Likewise, a dominant negative catalytically-inactive kinase which interacts normally with target proteins but does not phosphorylate the target proteins can reduce phosphorylation of the target proteins in response to a cellular signal. Similarly, a dominant negative transcription factor which binds to a promoter site in the control region of a gene but does not increase gene transcription can reduce the effect of a normal transcription factor by occupying promoter binding sites without increasing transcription.

The end result of the expression of a dominant negative polypeptide in a cell is a reduction in function of active proteins. One of ordinary skill in the art can assess the

potential for a dominant negative variant of a protein, and using standard mutagenesis techniques to create one or more dominant negative variant polypeptides. For example, given the teachings contained herein of CT antigens, especially those which are similar to known proteins which have known activities, one of ordinary skill in the art can modify the sequence of the CT antigens by site-specific mutagenesis, scanning mutagenesis, partial gene deletion or truncation, and the like. See, e.g., U.S. Patent No. 5,580,723 and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. The skilled artisan then can test the population of mutagenized polypeptides for diminution in a selected and/or for retention of such an activity. Other similar methods for creating and testing dominant negative variants of a protein will be apparent to one of ordinary skill in the art.

The invention as described herein has a number of uses, some of which are described elsewhere herein. First, the invention permits isolation of the CT antigen protein molecules. A variety of methodologies well-known to the skilled practitioner can be utilized to obtain isolated CT antigen molecules. The polypeptide may be purified from cells which naturally produce the polypeptide by chromatographic means or immunological recognition. Alternatively, an expression vector may be introduced into cells to cause production of the polypeptide. In another method, mRNA transcripts may be microinjected or otherwise introduced into cells to cause production of the encoded polypeptide. Translation of mRNA in cell-free extracts such as the reticulocyte lysate system also may be used to produce polypeptide. Those skilled in the art also can readily follow known methods for isolating CT antigen polypeptides. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography and immune-affinity chromatography.

The invention also makes it possible to isolate proteins which bind to CT antigens as disclosed herein, including antibodies and cellular binding partners of the CT antigens. Additional uses are described further herein.

The isolation and identification of CT antigen genes also makes it possible for the artisan to diagnose a disorder characterized by expression of CT antigens. These methods involve determining expression of one or more CT antigen nucleic acids, and/or encoded CT antigen polypeptides and/or peptides derived therefrom. In the former situation, such determinations can be carried out via any standard nucleic acid determination assay, including the polymerase chain reaction, or assaying with labeled hybridization probes. In

the latter two situations, such determinations can be carried out by immunoassays including, for example, ELISAs for the CT antigens, immunohistochemistry on tissue samples, and screening patient antisera for recognition of the polypeptide.

The invention further includes nucleic acid or protein microarrays with CT antigens or nucleic acids encoding such polypeptides. In this aspect of the invention, standard techniques of microarray technology are utilized to assess expression of the CT antigens and/or identify biological constituents that bind such polypeptides. The constituents of biological samples include antibodies, lymphocytes (particularly T lymphocytes), and the like. Protein microarray technology, which is also known by other names including: protein chip technology and solid-phase protein array technology, is well known to those of ordinary skill in the art and is based on, but not limited to, obtaining an array of identified peptides or proteins on a fixed substrate, binding target molecules or biological constituents to the peptides, and evaluating such binding. See, e.g., G. MacBeath and S.L. Schreiber, "Printing Proteins as Microarrays for High-Throughput Function Determination," *Science* 289(5485):1760-1763, 2000. Nucleic acid arrays, particularly arrays that bind CT antigens, also can be used for diagnostic applications, such as for identifying subjects that have a condition characterized by CT antigen expression.

Microarray substrates include but are not limited to glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. The microarray substrates may be coated with a compound to enhance synthesis of a probe (peptide or nucleic acid) on the substrate. Coupling agents or groups on the substrate can be used to covalently link the first nucleotide or amino acid to the substrate. A variety of coupling agents or groups are known to those of skill in the art. Peptide or nucleic acid probes thus can be synthesized directly on the substrate in a predetermined grid. Alternatively, peptide or nucleic acid probes can be spotted on the substrate, and in such cases the substrate may be coated with a compound to enhance binding of the probe to the substrate. In these embodiments, presynthesized probes are applied to the substrate in a precise, predetermined volume and grid pattern, preferably utilizing a computer-controlled robot to apply probe to the substrate in a contact-printing manner or in a non-contact manner such as ink jet or piezo-electric delivery. Probes may be covalently linked to the substrate.

Targets are peptides or proteins and may be natural or synthetic. The tissue may be obtained from a subject or may be grown in culture (e.g. from a cell line).

In some embodiments of the invention one or more control peptide or protein molecules are attached to the substrate. Preferably, control peptide or protein molecules allow determination of factors such as peptide or protein quality and binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success.

In other embodiments, one or more control peptide or nucleic acid molecules are attached to the substrate. Preferably, control nucleic acid molecules allow determination of factors such as binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success.

Nucleic acid microarray technology, which is also known by other names including: DNA chip technology, gene chip technology, and solid-phase nucleic acid array technology, is well known to those of ordinary skill in the art and is based on, but not limited to, obtaining an array of identified nucleic acid probes on a fixed substrate, labeling target molecules with reporter molecules (e.g., radioactive, chemiluminescent, or fluorescent tags such as fluorescein, Cye3-dUTP, or Cye5-dUTP), hybridizing target nucleic acids to the probes, and evaluating target-probe hybridization. A probe with a nucleic acid sequence that perfectly matches the target sequence will, in general, result in detection of a stronger reporter-molecule signal than will probes with less perfect matches. Many components and techniques utilized in nucleic acid microarray technology are presented in *The Chipping Forecast*, Nature Genetics, Vol.21, Jan 1999, the entire contents of which is incorporated by reference herein.

According to the present invention, nucleic acid microarray substrates may include but are not limited to glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. In all embodiments a glass substrate is preferred. According to the invention, probes are selected from the group of nucleic acids including, but not limited to: DNA, genomic DNA, cDNA, and oligonucleotides; and may be natural or synthetic. Oligonucleotide probes preferably are 20 to 25-mer oligonucleotides and DNA/cDNA probes preferably are 500 to 5000 bases in length, although other lengths may be used. Appropriate probe length may be determined by one of ordinary skill in the art by following art-known procedures. In one embodiment, preferred probes are sets of two or more of the CT antigen nucleic acid molecules set forth herein. Probes may be purified to remove contaminants using standard methods known to those of ordinary skill in the art such as gel filtration or precipitation.

In one embodiment, the microarray substrate may be coated with a compound to enhance synthesis of the probe on the substrate. Such compounds include, but are not limited to, oligoethylene glycols. In another embodiment, coupling agents or groups on the substrate can be used to covalently link the first nucleotide or oligonucleotide to the substrate. These agents or groups may include, for example, amino, hydroxy, bromo, and carboxy groups. These reactive groups are preferably attached to the substrate through a hydrocarbyl radical such as an alkylene or phenylene divalent radical, one valence position occupied by the chain bonding and the remaining attached to the reactive groups. These hydrocarbyl groups may contain up to about ten carbon atoms, preferably up to about six carbon atoms. Alkylene radicals are usually preferred containing two to four carbon atoms in the principal chain. These and additional details of the process are disclosed, for example, in U.S. Patent 4,458,066, which is incorporated by reference in its entirety.

In one embodiment, probes are synthesized directly on the substrate in a predetermined grid pattern using methods such as light-directed chemical synthesis, photochemical deprotection, or delivery of nucleotide precursors to the substrate and subsequent probe production.

In another embodiment, the substrate may be coated with a compound to enhance binding of the probe to the substrate. Such compounds include, but are not limited to: polylysine, amino silanes, amino-reactive silanes (Chipping Forecast, 1999) or chromium. In this embodiment, presynthesized probes are applied to the substrate in a precise, predetermined volume and grid pattern, utilizing a computer-controlled robot to apply probe to the substrate in a contact-printing manner or in a non-contact manner such as ink jet or piezo-electric delivery. Probes may be covalently linked to the substrate with methods that include, but are not limited to, UV-irradiation. In another embodiment probes are linked to the substrate with heat.

Targets for microarrays are nucleic acids selected from the group, including but not limited to: DNA, genomic DNA, cDNA, RNA, mRNA and may be natural or synthetic. In all embodiments, nucleic acid target molecules from human tissue are preferred. The tissue may be obtained from a subject or may be grown in culture (e.g. from a cell line).

In embodiments of the invention one or more control nucleic acid molecules are attached to the substrate. Preferably, control nucleic acid molecules allow determination of factors such as nucleic acid quality and binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success. Control nucleic

acids may include but are not limited to expression products of genes such as housekeeping genes or fragments thereof.

In some embodiments, one or more control peptide or nucleic acid molecules are attached to the substrate. Preferably, control nucleic acid molecules allow determination of factors such as binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success.

Expression of CT antigen polypeptides can also be determined using protein measurement methods. Preferred methods of specifically and quantitatively measuring proteins include, but are not limited to: mass spectroscopy-based methods such as surface enhanced laser desorption ionization (SELDI; e.g., Ciphergen ProteinChip System, Ciphergen Biosystems, Fremont CA), non-mass spectroscopy-based methods, and immunohistochemistry-based methods such as two-dimensional gel electrophoresis.

SELDI methodology may, through procedures known to those of ordinary skill in the art, be used to vaporize microscopic amounts of tumor protein and to create a "fingerprint" of individual proteins, thereby allowing simultaneous measurement of the abundance of many proteins in a single sample. Preferably SELDI-based assays may be utilized to classify tumor samples with respect to the expression of a variety of CT antigens. Such assays preferably include, but are not limited to the following examples. Gene products discovered by RNA microarrays may be selectively measured by specific (antibody mediated) capture to the SELDI protein disc (e.g., selective SELDI). Gene products discovered by protein screening (e.g., with 2-D gels), may be resolved by "total protein SELDI" optimized to visualize those particular markers of interest from among CT antigens.

Tumors can be classified based on the measurement of multiple CT antigens. Classification based on CT antigen expression can be used to stage disease, monitor progression or regression of disease, and select treatment strategies for the cancer patients.

The invention also involves agents such as polypeptides which bind to CT antigen polypeptides. Such binding agents can be used, for example, in screening assays to detect the presence or absence of CT antigen polypeptides and complexes of CT antigen polypeptides and their binding partners and in purification protocols to isolated CT antigen polypeptides and complexes of CT antigen polypeptides and their binding partners. Such agents also can be used to inhibit the native activity of the CT antigen polypeptides, for example, by binding to such polypeptides.

The invention, therefore, embraces peptide binding agents which, for example, can be antibodies or fragments of antibodies having the ability to selectively bind to CT antigen polypeptides. Antibodies include polyclonal and monoclonal antibodies, prepared according to conventional methodology.

Significantly, as is well-known in the art, only a small portion of an antibody molecule, the paratope, is involved in the binding of the antibody to its epitope (see, in general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology Wiley & Sons, Inc., New York; Roitt, I. (1991) Essential Immunology, 7th Ed., Blackwell Scientific Publications, Oxford). The pFc' and Fc regions, for example, are effectors of the complement cascade but are not involved in antigen binding. An antibody from which the pFc' region has been enzymatically cleaved, or which has been produced without the pFc' region, designated an $F(ab')_2$ fragment, retains both of the antigen binding sites of an intact antibody. Similarly, an antibody from which the Fc region has been enzymatically cleaved, or which has been produced without the Fc region, designated an Fab fragment, retains one of the antigen binding sites of an intact antibody molecule. Proceeding further, Fab fragments consist of a covalently bound antibody light chain and a portion of the antibody heavy chain denoted Fd. The Fd fragments are the major determinant of antibody specificity (a single Fd fragment may be associated with up to ten different light chains without altering antibody specificity) and Fd fragments retain epitope-binding ability in isolation.

Within the antigen-binding portion of an antibody, as is well-known in the art, there are complementarity determining regions (CDRs), which directly interact with the epitope of the antigen, and framework regions (FRs), which maintain the tertiary structure of the paratope (see, in general, Clark, 1986; Roitt, 1991). In both the heavy chain Fd fragment and the light chain of IgG immunoglobulins, there are four framework regions (FR1 through FR4) separated respectively by three complementarity determining regions (CDR1 through CDR3). The CDRs, and in particular the CDR3 regions, and more particularly the heavy chain CDR3, are largely responsible for antibody specificity.

It is now well-established in the art that the non-CDR regions of a mammalian antibody may be replaced with similar regions of conspecific or heterospecific antibodies while retaining the epitopic specificity of the original antibody. This is most clearly manifested in the development and use of "humanized" antibodies in which non-human CDRs are covalently joined to human FR and/or Fc/pFc' regions to produce a functional antibody. See, e.g., U.S. patents 4,816,567, 5,225,539, 5,585,089, 5,693,762 and 5,859,205.

Fully human monoclonal antibodies also can be prepared by immunizing mice transgenic for large portions of human immunoglobulin heavy and light chain loci. See, e.g., U.S. patents 5,545,806, 6,150,584, and references cited therein. Following immunization of these mice (e.g., XenoMouse (Abgenix), HuMAb mice (Medarex/GenPharm)), monoclonal antibodies can be prepared according to standard hybridoma technology. These monoclonal antibodies will have human immunoglobulin amino acid sequences and therefore will not provoke human anti-mouse antibody (HAMA) responses when administered to humans.

Thus, as will be apparent to one of ordinary skill in the art, the present invention also provides for $F(ab')_2$, Fab, Fv and Fd fragments; chimeric antibodies in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric $F(ab')_2$ fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric Fab fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; and chimeric Fd fragment antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced by homologous human or non-human sequences. The present invention also includes so-called single chain antibodies.

Accordingly, the invention involves polypeptides of numerous size and type that bind specifically to CT antigen polypeptides, and complexes of both CT antigen polypeptides and their binding partners. These polypeptides may be derived also from sources other than antibody technology. For example, such polypeptide binding agents can be provided by degenerate peptide libraries which can be readily prepared in solution, in immobilized form or as phage display libraries. Combinatorial libraries also can be synthesized of peptides containing one or more amino acids. Libraries further can be synthesized of peptoids and non-peptide synthetic moieties.

Phage display can be particularly effective in identifying binding peptides useful according to the invention. Briefly, one prepares a phage library (using e.g. m13, fd, or lambda phage), displaying inserts from 4 to about 80 amino acid residues using conventional procedures. The inserts may represent, for example, a completely degenerate or biased array. One then can select phage-bearing inserts which bind to the CT antigen polypeptide. This process can be repeated through several cycles of reselection of phage that bind to the CT antigen polypeptide. Repeated rounds lead to enrichment of phage bearing particular sequences. DNA sequence analysis can be conducted to identify the sequences of the

expressed polypeptides. The minimal linear portion of the sequence that binds to the CT antigen polypeptide can be determined. One can repeat the procedure using a biased library containing inserts containing part or all of the minimal linear portion plus one or more additional degenerate residues upstream or downstream thereof. Yeast two-hybrid screening methods also may be used to identify polypeptides that bind to the CT antigen polypeptides. Thus, the CT antigen polypeptides of the invention, or a fragment thereof, can be used to screen peptide libraries, including phage display libraries, to identify and select peptide binding partners of the CT antigen polypeptides of the invention. Such molecules can be used, as described, for screening assays, for purification protocols, for interfering directly with the functioning of CT antigen and for other purposes that will be apparent to those of ordinary skill in the art.

As detailed herein, the foregoing antibodies and other binding molecules may be used for example to identify tissues expressing protein or to purify protein. Antibodies also may be coupled to specific diagnostic labeling agents for imaging of cells and tissues that express CT antigens or to therapeutically useful agents according to standard coupling procedures. Diagnostic agents include, but are not limited to, barium sulfate, iocetamic acid, iopanoic acid, ipodate calcium, diatrizoate sodium, diatrizoate meglumine, metrizamide, tyropanoate sodium and radiodiagnostics including positron emitters such as fluorine-18 and carbon-11, gamma emitters such as iodine-123, technitium-99m, iodine-131 and indium-111, nuclides for nuclear magnetic resonance such as fluorine and gadolinium. Other diagnostic agents useful in the invention will be apparent to one of ordinary skill in the art.

As used herein, "therapeutically useful agents" include any therapeutic molecule which desirably is targeted selectively to a cell expressing one of the cancer antigens disclosed herein, including antineoplastic agents, radioiodinated compounds, toxins, other cytostatic or cytolytic drugs, and so forth. Antineoplastic therapeutics are well known and include: aminoglutethimide, azathioprine, bleomycin sulfate, busulfan, carmustine, chlorambucil, cisplatin, cyclophosphamide, cyclosporine, cytarabidine, dacarbazine, dactinomycin, daunorubicin, doxorubicin, taxol, etoposide, fluorouracil, interferon-$\alpha$, lomustine, mercaptopurine, methotrexate, mitotane, procarbazine HCl, thioguanine, vinblastine sulfate and vincristine sulfate. Additional antineoplastic agents include those disclosed in Chapter 52, Antineoplastic Agents (Paul Calabresi and Bruce A. Chabner), and the introduction thereto, 1202-1263, of Goodman and Gilman's "The Pharmacological Basis of Therapeutics", Eighth Edition, 1990, McGraw-Hill, Inc. (Health Professions Division).

Toxins can be proteins such as, for example, pokeweed anti-viral protein, cholera toxin, pertussis toxin, ricin, gelonin, abrin, diphtheria exotoxin, or *Pseudomonas* exotoxin. Toxin moieties can also be high energy-emitting radionuclides such as cobalt-60.

In some embodiments, antibodies prepared according to the invention are specific for complexes of MHC molecules and the CT antigens described herein.

When "disorder" is used herein, it refers to any pathological condition where the CT antigens are expressed. An example of such a disorder is cancer, including but not limited to: biliary tract cancer; bladder cancer; breast cancer; brain cancer including glioblastomas and medulloblastomas; cervical cancer; choriocarcinoma; colon cancer; endometrial cancer; esophageal cancer; gastric cancer; head and neck cancer; hematological neoplasms including acute lymphocytic and myelogenous leukemia, multiple myeloma, AIDS-associated leukemias and adult T-cell leukemia lymphoma; intraepithelial neoplasms including Bowen's disease and Paget's disease; liver cancer; lung cancer including small cell lung cancer and non-small cell lung cancer; lymphomas including Hodgkin's disease and lymphocytic lymphomas; neuroblastomas; oral cancer including squamous cell carcinoma; ovarian cancer including those arising from epithelial cells, stromal cells, germ cells and mesenchymal cells; pancreatic cancer; prostate cancer; rectal cancer; sarcomas including leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, synovial sarcoma and osteosarcoma; skin cancer including melanoma, Kaposi's sarcoma, basocellular cancer, and squamous cell cancer; testicular cancer including germinal tumors such as seminoma, non-seminoma (teratomas, choriocarcinomas), stromal tumors, and germ cell tumors; thyroid cancer including thyroid adenocarcinoma and medullar carcinoma; transitional cancer and renal cancer including adenocarcinoma and Wilms tumor.

Samples of tissue and/or cells for use in the various methods described herein can be obtained through standard methods such as tissue biopsy, including punch biopsy and cell scraping, and collection of blood or other bodily fluids by aspiration or other methods.

In certain embodiments of the invention, an immunoreactive cell sample is removed from a subject. By "immunoreactive cell" is meant a cell which can mature into an immune cell (such as a B cell, a helper T cell, or a cytolytic T cell) upon appropriate stimulation. Thus immunoreactive cells include $CD34^+$ hematopoietic stem cells, immature T cells and immature B cells. When it is desired to produce cytolytic T cells which recognize a CT antigen, the immunoreactive cell is contacted with a cell which expresses a CT antigen under conditions favoring production, differentiation and/or selection of cytolytic T cells; the

differentiation of the T cell precursor into a cytolytic T cell upon exposure to antigen is similar to clonal selection of the immune system.

Some therapeutic approaches based upon the disclosure are premised on a response by a subject's immune system, leading to lysis of antigen presenting cells, such as cancer cells which present one or more CT antigens. One such approach is the administration of autologous CTLs specific to a CT antigen/MHC complex to a subject with abnormal cells of the phenotype at issue. It is within the ability of one of ordinary skill in the art to develop such CTLs *in vitro*. An example of a method for T cell differentiation is presented in International Application number PCT/US96/05607. Generally, a sample of cells taken from a subject, such as blood cells, are contacted with a cell presenting the complex and capable of provoking CTLs to proliferate. The target cell can be a transfectant, such as a COS cell. These transfectants present the desired complex of their surface and, when combined with a CTL of interest, stimulate its proliferation. COS cells are widely available, as are other suitable host cells. Specific production of CTL clones is well known in the art. The clonally expanded autologous CTLs then are administered to the subject.

Another method for selecting antigen-specific CTL clones has recently been described (Altman et al., *Science* 274:94-96, 1996; Dunbar et al., *Curr. Biol.* 8:413-416, 1998), in which fluorogenic tetramers of MHC class I molecule/peptide complexes are used to detect specific CTL clones. Briefly, soluble MHC class I molecules are folded *in vitro* in the presence of $\beta_2$-microglobulin and a peptide antigen which binds the class I molecule. After purification, the MHC/peptide complex is purified and labeled with biotin. Tetramers are formed by mixing the biotinylated peptide-MHC complex with labeled avidin (e.g. phycoerythrin) at a molar ratio or 4:1. Tetramers are then contacted with a source of CTLs such as peripheral blood or lymph node. The tetramers bind CTLs which recognize the peptide antigen/MHC class I complex. Cells bound by the tetramers can be sorted by fluorescence activated cell sorting to isolate the reactive CTLs. The isolated CTLs then can be expanded *in vitro* for use as described herein.

To detail a therapeutic methodology, referred to as adoptive transfer (Greenberg, *J. Immunol.* 136(5): 1917, 1986; Riddel et al., *Science* 257: 238, 1992; Lynch et al, *Eur. J. Immunol.* 21: 1403-1410,1991; Kast et al., *Cell* 59: 603-614, 1989), cells presenting the desired complex (e.g., dendritic cells) are combined with CTLs leading to proliferation of the CTLs specific thereto. The proliferated CTLs are then administered to a subject with a cellular abnormality which is characterized by certain of the abnormal cells presenting the

particular complex. The CTLs then lyse the abnormal cells, thereby achieving the desired therapeutic goal.

The foregoing therapy assumes that at least some of the subject's abnormal cells present the relevant HLA/CT antigen complex. This can be determined very easily, as the art is very familiar with methods for identifying cells which present a particular HLA molecule, as well as how to identify cells expressing DNA of the pertinent sequences, in this case a CT antigen sequence. Once cells presenting the relevant complex are identified via the foregoing screening methodology, they can be combined with a sample from a patient, where the sample contains CTLs. If the complex presenting cells are lysed by the mixed CTL sample, then it can be assumed that a CT antigen is being presented, and the subject is an appropriate candidate for the therapeutic approaches set forth *supra*.

Adoptive transfer is not the only form of therapy that is available in accordance with the invention. CTLs can also be provoked *in vivo*, using a number of approaches. One approach is the use of non-proliferative cells expressing the complex. The cells used in this approach may be those that normally express the complex, such as irradiated tumor cells or cells transfected with one or both of the genes necessary for presentation of the complex (i.e. the antigenic peptide and the presenting HLA molecule). Chen et al. (*Proc. Natl. Acad. Sci. USA* 88: 110-114,1991) exemplifies this approach, showing the use of transfected cells expressing HPVE7 peptides in a therapeutic regime. Various cell types may be used. Similarly, vectors carrying one or both of the genes of interest may be used. Viral or bacterial vectors are especially preferred. For example, nucleic acids which encode a CT antigen polypeptide or peptide may be operably linked to promoter and enhancer sequences which direct expression of the CT antigen polypeptide or peptide in certain tissues or cell types. The nucleic acid may be incorporated into an expression vector. Expression vectors may be unmodified extrachromosomal nucleic acids, plasmids or viral genomes constructed or modified to enable insertion of exogenous nucleic acids, such as those encoding CT antigen, as described elsewhere herein. Nucleic acids encoding a CT antigen also may be inserted into a retroviral genome, thereby facilitating integration of the nucleic acid into the genome of the target tissue or cell type. In these systems, the gene of interest is carried by a microorganism, e.g., a Vaccinia virus, pox virus, herpes simplex virus, retrovirus or adenovirus, and the materials de facto "infect" host cells. The cells which result present the complex of interest, and are recognized by autologous CTLs, which then proliferate.

A similar effect can be achieved by combining the CT antigen or an immune response stimulatory fragment thereof with an adjuvant to facilitate incorporation into antigen presenting cells *in vivo*. The CT antigen polypeptide is processed to yield the peptide partner of the HLA molecule while a CT antigen peptide may be presented without the need for further processing. Generally, subjects can receive an intradermal injection of an effective amount of the CT antigen. Initial doses can be followed by booster doses, following immunization protocols standard in the art.

The invention involves the use of various materials disclosed herein to "immunize" subjects or as "vaccines". As used herein, "immunization" or "vaccination" means increasing or activating an immune response against an antigen. It does not require elimination or eradication of a condition but rather contemplates the clinically favorable enhancement of an immune response toward an antigen. Generally accepted animal models can be used for testing of immunization against cancer using a CT antigen nucleic acid. For example, human cancer cells can be introduced into a mouse to create a tumor, and one or more CT antigen nucleic acids can be delivered by the methods described herein. The effect on the cancer cells (e.g., reduction of tumor size) can be assessed as a measure of the effectiveness of the CT antigen nucleic acid immunization. Of course, testing of the foregoing animal model using more conventional methods for immunization include the administration of one or more CT antigen polypeptides or peptides derived therefrom, optionally combined with one or more adjuvants and/or cytokines to boost the immune response. Methods for immunization, including formulation of a vaccine composition and selection of doses, route of administration and the schedule of administration (e.g. primary and one or more booster doses), are well known in the art. The tests also can be performed in humans, where the end point is to test for the presence of enhanced levels of circulating CTLs against cells bearing the antigen, to test for levels of circulating antibodies against the antigen, to test for the presence of cells expressing the antigen and so forth.

As part of the immunization compositions, one or more CT antigens or stimulatory fragments thereof are administered with one or more adjuvants to induce an immune response or to increase an immune response. An adjuvant is a substance incorporated into or administered with antigen which potentiates the immune response. Adjuvants may enhance the immunological response by providing a reservoir of antigen (extracellularly or within macrophages), activating macrophages and stimulating specific sets of lymphocytes. Adjuvants of many kinds are well known in the art. Specific examples of adjuvants include

monophosphoryl lipid A (MPL, SmithKline Beecham), a congener obtained after purification and acid hydrolysis of *Salmonella minnesota* Re 595 lipopolysaccharide; saponins including QS21 (SmithKline Beecham), a pure QA-21 saponin purified from *Quillja saponaria* extract; DQS21, described in PCT application WO96/33739 (SmithKline Beecham); QS-7, QS-17, QS-18, and QS-L1 (So et al., *Mol. Cells* 7:178-186, 1997); incomplete Freund's adjuvant; complete Freund's adjuvant; montanide; immunostimulatory oligonucleotides (see e.g. CpG oligonucleotides described by Kreig et al., *Nature* 374:546-9, 1995); vitamin E and various water-in-oil emulsions prepared from biodegradable oils such as squalene and/or tocopherol. Preferably, the peptides are administered mixed with a combination of DQS21/MPL. The ratio of DQS21 to MPL typically will be about 1:10 to 10:1, preferably about 1:5 to 5:1 and more preferably about 1:1. Typically for human administration, DQS21 and MPL will be present in a vaccine formulation in the range of about 1 μg to about 100 μg. Other adjuvants are known in the art and can be used in the invention (*see, e.g.* Goding, *Monoclonal Antibodies: Principles and Practice*, 2nd Ed., 1986). Methods for the preparation of mixtures or emulsions of peptide and adjuvant are well known to those of skill in the art of vaccination.

Other agents which stimulate the immune response of the subject can also be administered to the subject. For example, other cytokines are also useful in vaccination protocols as a result of their lymphocyte regulatory properties. Many other cytokines useful for such purposes will be known to one of ordinary skill in the art, including interleukin-12 (IL-12) which has been shown to enhance the protective effects of vaccines (*see, e.g., Science* 268: 1432-1434, 1995), GM-CSF and IL-18. Thus cytokines can be administered in conjunction with antigens and adjuvants to increase the immune response to the antigens.

There are a number of immune response potentiating compounds that can be used in vaccination protocols. These include costimulatory molecules provided in either protein or nucleic acid form. Such costimulatory molecules include the B7-1 and B7-2 (CD80 and CD86 respectively) molecules which are expressed on dendritic cells (DC) and interact with the CD28 molecule expressed on the T cell. This interaction provides costimulation (signal 2) to an antigen/MHC/TCR stimulated (signal 1) T cell, increasing T cell proliferation and effector function. B7 also interacts with CTLA4 (CD152) on T cells and studies involving CTLA4 and B7 ligands indicate that the B7-CTLA4 interaction can enhance antitumor immunity and CTL proliferation (Zheng P., et al. *Proc. Natl. Acad. Sci. USA* 95 (11):6284-6289 (1998)).

B7 typically is not expressed on tumor cells so they are not efficient antigen presenting cells (APCs) for T cells. Induction of B7 expression would enable the tumor cells to stimulate more efficiently CTL proliferation and effector function. A combination of B7/IL-6/IL-12 costimulation has been shown to induce IFN-gamma and a Th1 cytokine profile in the T cell population leading to further enhanced T cell activity (Gajewski et al., *J. Immunol,* 154:5637-5648 (1995)). Tumor cell transfection with B7 has been discussed in relation to *in vitro* CTL expansion for adoptive transfer immunotherapy by Wang et al., (*J. Immunol.*, 19:1-8 (1986)). Other delivery mechanisms for the B7 molecule would include nucleic acid (naked DNA) immunization (Kim J., et al. *Nat Biotechnol.*, 15:7:641-646 (1997)) and recombinant viruses such as adeno and pox (Wendtner et al., *Gene Ther.*, 4:7:726-735 (1997)). These systems are all amenable to the construction and use of expression cassettes for the coexpression of B7 with other molecules of choice such as the antigens or fragment(s) of antigens discussed herein (including polytopes) or cytokines. These delivery systems can be used for induction of the appropriate molecules *in vitro* and for *in vivo* vaccination situations. The use of anti-CD28 antibodies to directly stimulate T cells *in vitro* and *in vivo* could also be considered. Similarly, the inducible co-stimulatory molecule ICOS which induces T cell responses to foreign antigen could be modulated, for example, by use of anti-ICOS antibodies (Hutloff et al., *Nature* 397:263-266, 1999).

Lymphocyte function associated antigen-3 (LFA-3) is expressed on APCs and some tumor cells and interacts with CD2 expressed on T cells. This interaction induces T cell IL-2 and IFN-gamma production and can thus complement but not substitute, the B7/CD28 costimulatory interaction (Parra et al., *J. Immunol.*, 158:637-642 (1997), Fenton et al., *J. Immunother.*, 21:2:95-108 (1998)).

Lymphocyte function associated antigen-1 (LFA-1) is expressed on leukocytes and interacts with ICAM-1 expressed on APCs and some tumor cells. This interaction induces T cell IL-2 and IFN-gamma production and can thus complement but not substitute, the B7/CD28 costimulatory interaction (Fenton et al., *J. Immunother.*, 21:2:95-108 (1998)). LFA-1 is thus a further example of a costimulatory molecule that could be provided in a vaccination protocol in the various ways discussed above for B7.

Complete CTL activation and effector function requires Th cell help through the interaction between the Th cell CD40L (CD40 ligand) molecule and the CD40 molecule expressed by DCs (Ridge et al., *Nature*, 393:474 (1998), Bennett et al., *Nature*, 393:478 (1998), Schoenberger et al., *Nature*, 393:480 (1998)). This mechanism of this costimulatory

signal is likely to involve upregulation of B7 and associated IL-6/IL-12 production by the DC (APC). The CD40-CD40L interaction thus complements the signal 1 (antigen/MHC-TCR) and signal 2 (B7-CD28) interactions.

The use of anti-CD40 antibodies to stimulate DC cells directly, would be expected to enhance a response to tumor antigens which are normally encountered outside of a inflammatory context or are presented by non-professional APCs (tumor cells). In these situations Th help and B7 costimulation signals are not provided. This mechanism might be used in the context of antigen pulsed DC based therapies or in situations where Th epitopes have not been defined within known TRA precursors.

A CT antigen polypeptide, or a fragment thereof, also can be used to isolate their native binding partners. Isolation of such binding partners may be performed according to well-known methods. For example, isolated CT antigen polypeptides can be attached to a substrate (e.g., chromatographic media, such as polystyrene beads, or a filter), and then a solution suspected of containing the binding partner may be applied to the substrate. If a binding partner which can interact with CT antigen polypeptides is present in the solution, then it will bind to the substrate-bound CT antigen polypeptide. The binding partner then may be isolated.

It will also be recognized that the invention embraces the use of the CT antigen cDNA sequences in expression vectors, as well as to transfect host cells and cell lines, be these prokaryotic (e.g., *E. coli*), or eukaryotic (e.g., dendritic cells, B cells, CHO cells, COS cells, yeast expression systems and recombinant baculovirus expression in insect cells). Especially useful are mammalian cells such as human, mouse, hamster, pig, goat, primate, etc. They may be of a wide variety of tissue types, and include primary cells and cell lines. Specific examples include keratinocytes, peripheral blood leukocytes, bone marrow stem cells and embryonic stem cells. The expression vectors require that the pertinent sequence, i.e., those nucleic acids described *supra*, be operably linked to a promoter.

The invention also contemplates delivery of nucleic acids, polypeptides or peptides for vaccination. Delivery of polypeptides and peptides can be accomplished according to standard vaccination protocols which are well known in the art. In another embodiment, the delivery of nucleic acid is accomplished by *ex vivo* methods, i.e. by removing a cell from a subject, genetically engineering the cell to include a CT antigen, and reintroducing the engineered cell into the subject. One example of such a procedure is the use of dendritic cells as delivery and antigen presentation vehicles for the administration of CT antigens in vaccine

therapies. Another example of such a procedure is outlined in U.S. Patent 5,399,346 and in exhibits submitted in the file history of that patent, all of which are publicly available documents. In general, it involves introduction *in vitro* of a functional copy of a gene into a cell(s) of a subject, and returning the genetically engineered cell(s) to the subject. The functional copy of the gene is under operable control of regulatory elements which permit expression of the gene in the genetically engineered cell(s). Numerous transfection and transduction techniques as well as appropriate expression vectors are well known to those of ordinary skill in the art, some of which are described in PCT application WO95/00654. *In vivo* nucleic acid delivery using vectors such as viruses and targeted liposomes also is contemplated according to the invention.

In preferred embodiments, a virus vector for delivering a nucleic acid encoding a CT antigen is selected from the group consisting of adenoviruses, adeno-associated viruses, poxviruses including vaccinia viruses and attenuated poxviruses, Semliki Forest virus, Venezuelan equine encephalitis virus, retroviruses, Sindbis virus, and Ty virus-like particle. Examples of viruses and virus-like particles which have been used to deliver exogenous nucleic acids include: replication-defective adenoviruses (e.g., Xiang et al., *Virology* 219:220-227, 1996; Eloit et al., *J. Virol.* 7:5375-5381, 1997; Chengalvala et al., *Vaccine* 15:335-339, 1997), a modified retrovirus (Townsend et al., *J. Virol.* 71:3365-3374, 1997), a nonreplicating retrovirus (Irwin et al., *J. Virol.* 68:5036-5044, 1994), a replication defective Semliki Forest virus (Zhao et al., *Proc. Natl. Acad. Sci. USA* 92:3009-3013, 1995), canarypox virus and highly attenuated vaccinia virus derivative (Paoletti, *Proc. Natl. Acad. Sci. USA* 93:11349-11353, 1996), non-replicative vaccinia virus (Moss, *Proc. Natl. Acad. Sci. USA* 93:11341-11348, 1996), replicative vaccinia virus (Moss, *Dev. Biol. Stand.* 82:55-63, 1994), Venzuelan equine encephalitis virus (Davis et al., *J. Virol.* 70:3781-3787, 1996), Sindbis virus (Pugachev et al., *Virology* 212:587-594, 1995), and Ty virus-like particle (Allsopp et al., *Eur. J. Immunol* 26:1951-1959, 1996). In preferred embodiments, the virus vector is an adenovirus or an alphavirus.

Another preferred virus for certain applications is the adeno-associated virus, a double-stranded DNA virus. The adeno-associated virus is capable of infecting a wide range of cell types and species and can be engineered to be replication-deficient. It further has advantages, such as heat and lipid solvent stability, high transduction frequencies in cells of diverse lineages, including hematopoietic cells, and lack of superinfection inhibition thus allowing multiple series of transductions. The adeno-associated virus can integrate into

human cellular DNA in a site-specific manner, thereby minimizing the possibility of insertional mutagenesis and variability of inserted gene expression. In addition, wild-type adeno-associated virus infections have been followed in tissue culture for greater than 100 passages in the absence of selective pressure, implying that the adeno-associated virus genomic integration is a relatively stable event. The adeno-associated virus can also function in an extrachromosomal fashion.

In general, other preferred viral vectors are based on non-cytopathic eukaryotic viruses in which non-essential genes have been replaced with the gene of interest. Non-cytopathic viruses include retroviruses, the life cycle of which involves reverse transcription of genomic viral RNA into DNA with subsequent proviral integration into host cellular DNA. Adenoviruses and retroviruses have been approved for human gene therapy trials. In general, the retroviruses are replication-deficient (i.e., capable of directing synthesis of the desired proteins, but incapable of manufacturing an infectious particle). Such genetically altered retroviral expression vectors have general utility for the high-efficiency transduction of genes *in vivo*. Standard protocols for producing replication-deficient retroviruses (including the steps of incorporation of exogenous genetic material into a plasmid, transfection of a packaging cell lined with plasmid, production of recombinant retroviruses by the packaging cell line, collection of viral particles from tissue culture media, and infection of the target cells with viral particles) are provided in Kriegler, M., "Gene Transfer and Expression, A Laboratory Manual," W.H. Freeman Co., New York (1990) and Murry, E.J. Ed. "Methods in Molecular Biology," vol. 7, Humana Press, Inc., Cliffton, New Jersey (1991).

Preferably the foregoing nucleic acid delivery vectors: (1) contain exogenous genetic material that can be transcribed and translated in a mammalian cell and that can induce an immune response in a host, and (2) contain on a surface a ligand that selectively binds to a receptor on the surface of a target cell, such as a mammalian cell, and thereby gains entry to the target cell.

Various techniques may be employed for introducing nucleic acids of the invention into cells, depending on whether the nucleic acids are introduced *in vitro* or *in vivo* in a host. Such techniques include transfection of nucleic acid-$CaPO_4$ precipitates, transfection of nucleic acids associated with DEAE, transfection or infection with the foregoing viruses including the nucleic acid of interest, liposome mediated transfection, and the like. For certain uses, it is preferred to target the nucleic acid to particular cells. In such instances, a vehicle used for delivering a nucleic acid of the invention into a cell (e.g., a retrovirus, or

other virus; a liposome) can have a targeting molecule attached thereto. For example, a molecule such as an antibody specific for a surface membrane protein on the target cell or a ligand for a receptor on the target cell can be bound to or incorporated within the nucleic acid delivery vehicle. Preferred antibodies include antibodies which selectively bind a CT antigen, alone or as a complex with a MHC molecule. Especially preferred are monoclonal antibodies. Where liposomes are employed to deliver the nucleic acids of the invention, proteins which bind to a surface membrane protein associated with endocytosis may be incorporated into the liposome formulation for targeting and/or to facilitate uptake. Such proteins include capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half life, and the like. Polymeric delivery systems also have been used successfully to deliver nucleic acids into cells, as is known by those skilled in the art. Such systems even permit oral delivery of nucleic acids.

When administered, the therapeutic compositions of the present invention can be administered in pharmaceutically acceptable preparations. Such preparations may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, supplementary immune potentiating agents such as adjuvants and cytokines and optionally other therapeutic agents.

The therapeutics of the invention can be administered by any conventional route, including injection or by gradual infusion over time. The administration may, for example, be oral, intravenous, intraperitoneal, intramuscular, intracavity, subcutaneous, or transdermal. When antibodies are used therapeutically, a preferred route of administration is by pulmonary aerosol. Techniques for preparing aerosol delivery systems containing antibodies are well known to those of skill in the art. Generally, such systems should utilize components which will not significantly impair the biological properties of the antibodies, such as the paratope binding capacity (see, for example, Sciarra and Cutie, "Aerosols," in Remington's Pharmaceutical Sciences, 18th edition, 1990, pp. 1694-1712; incorporated by reference). Those of skill in the art can readily determine the various parameters and conditions for producing antibody aerosols without resort to undue experimentation. When using antisense preparations of the invention, slow intravenous administration is preferred.

The compositions of the invention are administered in effective amounts. An "effective amount" is that amount of a CT antigen composition that alone, or together with further doses, produces the desired response, e.g. increases an immune response to the CT

antigen. In the case of treating a particular disease or condition characterized by expression of one or more CT antigens, such as cancer, the desired response is inhibiting the progression of the disease. This may involve only slowing the progression of the disease temporarily, although more preferably, it involves halting the progression of the disease permanently. This can be monitored by routine methods or can be monitored according to diagnostic methods of the invention discussed herein. The desired response to treatment of the disease or condition also can be delaying the onset or even preventing the onset of the disease or condition.

Such amounts will depend, of course, on the particular condition being treated, the severity of the condition, the individual patient parameters including age, physical condition, size and weight, the duration of the treatment, the nature of concurrent therapy (if any), the specific route of administration and like factors within the knowledge and expertise of the health practitioner. These factors are well known to those of ordinary skill in the art and can be addressed with no more than routine experimentation. It is generally preferred that a maximum dose of the individual components or combinations thereof be used, that is, the highest safe dose according to sound medical judgment. It will be understood by those of ordinary skill in the art, however, that a patient may insist upon a lower dose or tolerable dose for medical reasons, psychological reasons or for virtually any other reasons.

The pharmaceutical compositions used in the foregoing methods preferably are sterile and contain an effective amount of CT antigen or nucleic acid encoding CT antigen for producing the desired response in a unit of weight or volume suitable for administration to a patient. The response can, for example, be measured by determining the immune response following administration of the CT antigen composition via a reporter system by measuring downstream effects such as gene expression, or by measuring the physiological effects of the CT antigen composition, such as regression of a tumor or decrease of disease symptoms. Other assays will be known to one of ordinary skill in the art and can be employed for measuring the level of the response.

The doses of CT antigen compositions (e.g., polypeptide, peptide, antibody, cell or nucleic acid) administered to a subject can be chosen in accordance with different parameters, in particular in accordance with the mode of administration used and the state of the subject. Other factors include the desired period of treatment. In the event that a response in a subject is insufficient at the initial doses applied, higher doses (or effectively higher doses by a

different, more localized delivery route) may be employed to the extent that patient tolerance permits.

In general, for treatments for eliciting or increasing an immune response, doses of CT antigen are formulated and administered in doses between 1 ng and 1 mg, and preferably between 10 ng and 100 μg, according to any standard procedure in the art. Where nucleic acids encoding CT antigen or variants thereof are employed, doses of between 1 ng and 0.1 mg generally will be formulated and administered according to standard procedures. Other protocols for the administration of CT antigen compositions will be known to one of ordinary skill in the art, in which the dose amount, schedule of injections, sites of injections, mode of administration (e.g., intra-tumoral) and the like vary from the foregoing. Administration of CT antigen compositions to mammals other than humans, e.g. for testing purposes or veterinary therapeutic purposes, is carried out under substantially the same conditions as described above.

Where CT antigen peptides are used for vaccination, modes of administration which effectively deliver the CT antigen and adjuvant, such that an immune response to the antigen is increased, can be used. For administration of a CT antigen peptide in adjuvant, preferred methods include intradermal, intravenous, intramuscular and subcutaneous administration. Although these are preferred embodiments, the invention is not limited by the particular modes of administration disclosed herein. Standard references in the art (e.g., *Remington's Pharmaceutical Sciences*, 18th edition, 1990) provide modes of administration and formulations for delivery of immunogens with adjuvant or in a non-adjuvant carrier.

When administered, the pharmaceutical preparations of the invention are applied in pharmaceutically-acceptable amounts and in pharmaceutically-acceptable compositions. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. Such preparations may routinely contain salts, buffering agents, preservatives, compatible carriers, and optionally other therapeutic agents. When used in medicine, the salts should be pharmaceutically acceptable, but non-pharmaceutically acceptable salts may conveniently be used to prepare pharmaceutically-acceptable salts thereof and are not excluded from the scope of the invention. Such pharmacologically and pharmaceutically-acceptable salts include, but are not limited to, those prepared from the following acids: hydrochloric, hydrobromic, sulfuric, nitric, phosphoric, maleic, acetic, salicylic, citric, formic, malonic, succinic, and the like.

Also, pharmaceutically-acceptable salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts.

A CT antigen composition may be combined, if desired, with a pharmaceutically-acceptable carrier. The term "pharmaceutically-acceptable carrier" as used herein means one or more compatible solid or liquid fillers, diluents or encapsulating substances which are suitable for administration into a human. The term "carrier" denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. The components of the pharmaceutical compositions also are capable of being co-mingled with the molecules of the present invention, and with each other, in a manner such that there is no interaction which would substantially impair the desired pharmaceutical efficacy.

The pharmaceutical compositions may contain suitable buffering agents, including: acetic acid in a salt; citric acid in a salt; boric acid in a salt; and phosphoric acid in a salt.

The pharmaceutical compositions also may contain, optionally, suitable preservatives, such as: benzalkonium chloride; chlorobutanol; parabens and thimerosal.

The pharmaceutical compositions may conveniently be presented in unit dosage form and may be prepared by any of the methods well-known in the art of pharmacy. All methods include the step of bringing the active agent into association with a carrier which constitutes one or more accessory ingredients. In general, the compositions are prepared by uniformly and intimately bringing the active compound into association with a liquid carrier, a finely divided solid carrier, or both, and then, if necessary, shaping the product.

Compositions suitable for oral administration may be presented as discrete units, such as capsules, tablets, lozenges, each containing a predetermined amount of the active compound. Other compositions include suspensions in aqueous liquids or non-aqueous liquids such as a syrup, elixir or an emulsion.

Compositions suitable for parenteral administration conveniently comprise a sterile aqueous or non-aqueous preparation of CT antigen polypeptides or nucleic acids, which is preferably isotonic with the blood of the recipient. This preparation may be formulated according to known methods using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation also may be a sterile injectable solution or suspension in a non-toxic parenterally-acceptable diluent or solvent, for example, as a solution in 1,3-butane diol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution. In addition,

sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil may be employed including synthetic mono-or di-glycerides. In addition, fatty acids such as oleic acid may be used in the preparation of injectables. Carrier formulation suitable for oral, subcutaneous, intravenous, intramuscular, etc. administrations can be found in *Remington's Pharmaceutical Sciences*, Mack Publishing Co., Easton, PA.

As used herein with respect to nucleic acids, the term "isolated" means: (i) amplified *in vitro* by, for example, polymerase chain reaction (PCR); (ii) recombinantly produced by cloning; (iii) purified, as by cleavage and gel separation; or (iv) synthesized by, for example, chemical synthesis. An isolated nucleic acid is one which is readily manipulable by recombinant DNA techniques well known in the art. Thus, a nucleotide sequence contained in a vector in which 5' and 3' restriction sites are known or for which polymerase chain reaction (PCR) primer sequences have been disclosed is considered isolated but a nucleic acid sequence existing in its native state in its natural host is not. An isolated nucleic acid may be substantially purified, but need not be. For example, a nucleic acid that is isolated within a cloning or expression vector is not pure in that it may comprise only a tiny percentage of the material in the cell in which it resides. Such a nucleic acid is isolated, however, as the term is used herein because it is readily manipulable by standard techniques known to those of ordinary skill in the art. An isolated nucleic acid as used herein is not a naturally occurring chromosome.

As used herein with respect to polypeptides, "isolated" means separated from its native environment and present in sufficient quantity to permit its identification or use. Isolated, when referring to a protein or polypeptide, means, for example: (i) selectively produced by expression cloning or (ii) purified as by chromatography or electrophoresis. Isolated proteins or polypeptides may, but need not be, substantially pure. The term "substantially pure" means that the proteins or polypeptides are essentially free of other substances with which they may be found in nature or *in vivo* systems to an extent practical and appropriate for their intended use. Substantially pure polypeptides may be produced by techniques well known in the art. Because an isolated protein may be admixed with a pharmaceutically acceptable carrier in a pharmaceutical preparation, the protein may comprise only a small percentage by weight of the preparation. The protein is nonetheless isolated in that it has been separated from the substances with which it may be associated in living systems, i.e. isolated from other proteins.

## Examples

**Example 1: Identification of CT antigens**

Much attention has been given to the potential of CT antigens as targets for cancer vaccine development, and, other than mutational antigens and virus encoded antigens, they clearly represent the most specific tumor antigens discovered to date. However, the CT antigens also provide a new way to think about cancer and its evolution during the course of the disease.

The starting point for this view is the fact that CT antigen expression is restricted to early germ cell development and cancer. Germ cells give rise to gametes (oocytes and spermatocytes) and trophoblastic cells that contribute to the formation of the chorion and the placenta. Primitive germ cells arise in the wall of the yolk sack and during embryogenesis migrate to the future site of the gonads. In oogenesis, the process begins before birth, with oogonia differentiating into primary oocytes. The primary oocytes, which reach their maximal numbers during fetal development, are arrested at the initial phase of meiosis, and do not renew and complete meiosis until ovulation and fertilization. In contrast, spermatogenesis begins at puberty and is a continuous process of mitosis to maintain the spermatogonia pool and meiosis to generate the mature sperm population. CT antigens, like SCP-1 and OY-TES-1, the proacrosomal binding protein precursor, are clearly important in gametogenesis, and it is likely that the other CT antigens with their restricted expression in gametes and trophoblasts also play a critical role in early germ cell development.

One possibility to account for aberrant CT expression in cancer relates to the global demethylation associated with certain cancers (42). The promoter region of the MAGE gene has binding sites for transcriptional activators and these sites are methylated in normal somatic cells but demethylated in MAGE-expressing cancer cells and testis. Although cancer-associated demethylation could therefore account for CT (MAGE) expression in tumors, it does not easily accommodate the usual observation of non-coordinate expression patterns (sets) of different CT antigens in most tumors. Also, the marked heterogeneity in CT expression in some tumors (34, 43) is also not easily explicable by a global demethylation process.

Another mechanism for reactivating CT expression in cancer has to do with mutations in regulatory regions of the CT genes. Although no mutations in CT genes have been found to date, more extensive sequencing, particularly in the promoter region, needs to be done

before this can be excluded. However, mutation of CT genes is unlikely to be a common mechanism for the induction of CT expression in cancer.

Another possibility to account for the appearance of CT antigens in cancer is the induction or activation of a gametogenic program in cancer. According to this view, the different CT sets seen in cancer would replicate the corresponding sets of CT antigens normally expressed during different stages of gametogenesis or trophoblast development. Triggering events for inducing the gametogenic program could be a mutation in an as yet unidentified master switch in germ cell development, or an activation of this master switch by threshold mutations in oncogenes, suppressor genes, or other genes in cancer. It is also possible that activation of a single CT gene could be the switch for activating other genes in the gametogenic program. Supporting evidence for this idea comes from the study of synovial sarcoma, where a translocation event involving the SYT gene on chromosome 18 and the SSX-1 or SSX-2 gene on chromosome X is associated with high expression of unrelated CT antigens, such as NY-ESO-1 and MAGE (44, 45). Extending this line of reasoning and relating it to the role of demethylation in the appearance of CT antigens, a demethylation state in cancer (whatever its cause) could induce the gametogenic program and result in the activation of silent CT genes. Alternatively, demethylation may be an intrinsic part of the gametogenic program and therefore a consequence, not a cause, of switching on the gametogenic program and CT genes in cancer.

In addition to questions about mechanisms for reactivating CT antigen expression in cancer, another important issue is whether expression of these genes in the cancer cell contributes to its malignant behavior. The finding that gametes, trophoblasts and cancers share a battery of antigens restricted to these cell types suggests extending the search for other shared characteristics.

It was a similarity in the biological features of trophoblasts and cancer cells that prompted the Scottish embryologist John Beard at the turn of the last century to propose his trophoblastic theory of cancer (46, 47). In his view, cancers arise from germ cells that stray or are arrested in their trek to the gonads. Under the influence of carcinogenic stimuli, such cells undergo a conversion to malignant trophoblastic cells. These malignant trophoblastic cells take on features of the resident cell types in different organs, but the resulting cancers, no matter their site of origin or how distinct they appear morphologically, are of trophoblastic origin. Beard ascribed the invasive, destructive and metastatic features of cancer to functions normally displayed by trophoblastic cells, e.g., invasion of blood vessels, growth into the

uterine wall, and spread beyond the uterus. From a contemporary perspective, Beard's idea that cancers are derived from arrested germ cells seems incompatible with our growing knowledge of serological and molecular markers that distinguish different pathways of normal differentiation and their preservation in cancer. Beard's insight that trophoblasts and cancer cells share common features is better explained by the induction of a gametogenic program in resident cancer cells, rather than the derivation of cancer from an aberrant germ cell. The end result, however, would be the same - selected features of cells undergoing gametogenesis and trophoblast development being imposed on transformed somatic cells.

In addition to CT antigens, other features shared by germ cells and cancer are identified. For example, SCP-1, a critical element in the meiotic program, is expressed in non-germ cell cancers. The induction of a meiotic program in a somatic cell, normal or malignant, likely leads to chromosomal anarchy, a prime feature of advanced cancers. Accordingly, other proteins uniquely associated with meiosis and expressed in cancer cells also are identified as candidate CT antigens.

OY-TES-1, the proacrosin binding protein precursor that is part of the unique program leading to the formation of spermatozoa, has been identified as a CT antigen. Accordingly, other mature sperm-specific gene products that are expressed in cancer cells also are identified as candidate CT antigens.

In addition, expression of CT antigens by trophoblasts sheds new light on an old issue - the much studied sporadic production of human chorionic gonadotropin (HCG) and other trophoblastic hormones by human cancers (e.g., 48, 49, 50). The production of HCG by cancer cells has been generally viewed as yet another indication of the genetic instability of cancer cells, resulting in the random and aberrant activation of silent genes during carcinogenesis and tumor progression. However, it can also be viewed as a consequence of the induction of a gametogenic/trophoblastic program in cancer, one that would also result in the semi-coordinate expression of CT antigens. Activation of this program would also confer other properties of germ cells, gametes, and trophoblasts on cancer cells, but these are more difficult to relate in any precise fashion. Nonetheless, immortalization, invasion, lack of adhesion, migratory behavior, induction of blood vessels, demethylation, and downregulation of MHC, are some features shared by cancer and by cells undergoing germ cell/gamete/trophoblast differentiation pathways. The metastatic properties of cancer may also have counterparts in the migratory behavior of germ cells, and in the propensity of

normal trophoblast cells to migrate to other organs, such as the lung, during normal pregnancy, but then to undergo involution at term.

In pursing the idea of a program change in cancer leading to the expression of gametogenic features, a hypothesis termed "Gametogenic Program Induction in Cancer" (GPIC), it might be well to distinguish at least four different pathways involved in germ cell development: A) germ cell → germ cell, B) germ cell → oogonia → oocytes, C) germ cell → spermatogonia → sperm, and D) germ cell → trophoblast. The meiotic program would be common to B and C, proteins like OY-TES-1 would be restricted to C, and HCG would be a characteristic of D. The reason for distinguishing these pathways and ultimately stages in each pathway is that the variety of patterns or sets of CT antigens observed in different cancers may be a reflection of the germ cell program, e.g., pathway and stage that has been induced in these cancers.

With this background and framework of thinking about the relation of gametogenesis and cancer development, there are a number of approaches to be taken to identify additional CT antigens.

1. The search for new CT antigens is accomplished using several methodologies, including SEREX (see, for example, ref. 10), particularly with libraries from testis, normal or malignant trophoblasts, or tumors or tumor cell lines (growing with or without demethylating agents) that express a range of CT antigens, and by extending the use of representational difference analysis. Bioinformatics and chip technology are used for mining databanks for transcripts that show cancer/gamete/trophoblast specificity (e.g., screening annotation of sequence records).

2. The expression pattern of known CT antigens in normal gametogenesis and trophoblast development is determined to identify markers that distinguish different pathways and stages in the normal gametogenic program. This information provides a basis for interpreting the complex patterns of CT expression in cancers in relation to gametogenic pathways/stages, and provides new ways to classify cancer on the basis of CT phenotypes.

3. The frequency of expression of individual CT antigens in different tumor types has been defined for those CT antigens known to date. In addition to analyzing frequency of expression for CT antigens identified by the methods described herein, additional information is gathered about the composite CT phenotype of individual tumors, and how frequently these composite CT patterns are seen in tumors of different origin. Databases of clinical, genotypic, phenotypic and CT antigen expression data for individual tumors are established

to compare the properties of individual tumors and establish correlations between the data. With this information, correlations of CT expression with other biological features of the tumor, e.g., growth rate, local vs. invasive, primary vs. metastatic, different metastatic deposits in the same patient, etc. can be established.

4. Determining which stage in the life history of cancer that CT (gametogenic) features are induced can be approached in model systems in the mouse, *in vitro* systems with human cells, or with naturally occurring tumors in man that show incremental stages in tumor progression. As discussed above, there is evidence that CT expression is a sign of greater malignancy.

5. The heterogeneous expression of CT antigens in a large proportion of human cancers needs to be understood. This may reflect a quantitative difference in levels of mRNA/protein in $CT^+$ and $CT^-$ cells, or there may be a qualitative distinction between $CT^+$ and $CT^-$ cells in CT mRNA/protein expression. Laser dissection microscopy may be one way to analyze this question and cloning of tumor cells from a tumor with heterogeneous CT expression is another approach to understand heterogeneous expression. There is a growing impression that established human cancer cell lines show a higher frequency of CT antigen expression than what would be expected from CT typing of the corresponding tumor type, particularly tumors with a low frequency of CT expression. This could be a secondary consequence of *in vitro* culture, or it could be that $CT^+$ cells (even if they represent only a minority population of the tumor) have a growth advantage for propagating *in vitro*, and possibly also *in vivo*.

6. Although CT antigens provide a strong link between the gametogenic program and cancer, it is determined whether other distinguishing features of gamete development are expressed by cancer and whether their expression is correlated with CT antigen expression. The many reports over the last three decades of HCG production by certain human cancers provides a specific starting point to explore this issue and ask whether the production of HCG is correlated with CT antigen expression, particularly a unique pattern of CT expression, such as a pattern reflecting the trophoblast program.

7. Transgenic and knock-out approaches using mouse CT counterparts, and transfection analysis with CT coding genes in normal and malignant human cells are performed to define the role of CT antigens in gametogenesis and trophoblast development and their functional significance in cancer.

**Example 2: Identification of testis-specific gene as novel CT antigens expressed in multiple tumors**

**Materials and Methods**

***Sperm proteins***

A number of proteins have been identified as sperm-specific gene products in the literature. These include the proteins listed in Table 2. These are proteins involved in sperm-egg interaction, enzymes present in sperm, and others. SPAN-X was shown to be homologous to the known CT antigen CTp11 (17), and not analyzed in this study.

Table 2: Sperm Proteins

| Antigens | Species | Function/Characteristics |
|---|---|---|
| **Proteins involved in sperm-egg interaction** | | |
| • SP-10 | Human | Acrosomal antigen |
| • SP17 | Human, rabbit, mouse | Zona pellucida (ZP) binding in vitro |
| • NZ-1 | Mouse | ZP binding, tyrosine phosphorylation activity |
| • NZ-2 | Human | ZP binding, tyrosine phosphorylation activity |
| • FA-1 | Mouse | ZP binding, sperm capacitation |
| **Enzyme present in sperm** | | |
| • Acrosin | Human, mouse | Serine protease localized in sperm acrosome |
| • PH-20 | Guinea pig, human | Hyaluronidase activity, sperm penetration of the layer of cumulus cells surrounding oocyte |
| • LDH-$C_4$ | Mouse | Lactate dehydrogenase-$C_4$ |
| **Others** | | |
| • SP32 (OY-TES-1) | Human, mouse, guinea pig, pig | Proacrosin binding protein |
| • AKAP110 | Human, mouse | A-kinase anchoring protein |
| • ASP | Human | AKAP-associated protein |
| • Ropporin | Human | AKAP-associated protein |
| • CS-1 | Human | Cleavage signal protein |
| • SPAG9 (HSS) | Human | Sperm surface protein |
| • NYD-sp10 | Human | |
| • SPAN-X/CTp11 | Human | Nuclear protein |

***mRNA Isolation and cDNA Synthesis***

mRNA from malignant tissues was purified using the QuickPrep Micro mRNA Purification Kit (Amersham Pharmacia, Piscataway, NJ). mRNA was reverse transcribed into single strand cDNA using Moloney murine leukemia virus reverse transcriptase and oligo $(dT)_{15}$ as a primer (Amersham Pharmacia). cDNAs were tested for integrity by amplification of G3PDH transcripts in a 30 cycle reaction.

***Reverse Transcription-PCR (RT-PCR)***

To amplify cDNA segments from normal tissue (Multiple Tissue cDNA panel, CLONTECH, Palo Alto, CA) and malignant tissues, the primers for the respective genes were designed (Table 3). To avoid amplification of contaminating genomic DNA, primers were placed in different exons. RT-PCR was performed by using 30 amplification cycles and followed by a 10-min elongation step at 72°C. The PCR products were analyzed by agarose gel electrophoresis and capillary electrophoresis on a microtip device (DNA 7500 LabChip, Caliber Technologies, Mountain View, CA) by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and assessed for a single amplification product of the correct size.

***Real-time quantitative PCR***

A two-step real-time RT-PCR was used to determine relative expression levels of sperm protein mRNA using ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems, Foster City, CA). Primer pairs specific for *NY-ESO-1*, *OY-TES-1*, *SP17*, *acrosin*, *PH-20*, *AKAP110*, *ASP*, *CS*-1 and *SPAG9* used were listed in Table 3. For *SP-10*, *ropporin* and *NYD-sp10*, newly designed primer pairs were used: *SP-10*-5': 5'-CCAGAGGAACATCAAGTCAGC-3' (SEQ ID NO:11); *SP-10*-3': 5'-ATATTGTGCCTGTAGATGTG-3' (SEQ ID NO:12), product size 515bp; *ropporin*-5': 5'-TGCCGAAAATGCTGAAGGAG-3' (SEQ ID NO:13); *ropporin*-3': 5'-GTAGACAAACTGGAAGGTGC-3' (SEQ ID NO:14), product size 455bp; *NYD-sp10*-5': 5'-TACATTGAGTGGCTGGATAC-3' (SEQ ID NO:15); *NYD-sp10*-3': 5'-AGGTAGAGCACGTAGTCATC-3' (SEQ ID NO:16), product size 212bp. PCR was performed using SYBR Green PCR Core Reagent kit (Perkin-Elmer Applied Biosystems). The thermal cycling conditions comprised an initial denaturation step at 95°C for 10 min and 40 cycles at 95°C for 15 sec and 60°C for 1 min. The house keeping gene β-actin was used for internal normalization. Experiments were performed in duplicate for each data point.

Final results, expressed as n-fold differences in sperm protein gene expression relative to β-actin gene and normal testis (the calibrator) were determined in exponent as follows:

$$n = 2^{-(\Delta Ct\ sample - \Delta Ct\ calibrator)}$$

where ΔCt values of the sample and calibrator are determined by subtracting the average Ct value of the sperm gene from the average Ct value of the β-actin gene.

Table 3. Primer pairs used in this study

| Gene | Sequence of primer pair [1] | Annealing temperature (°C) | PCR product size (bp) | SEQ ID NO: |
|---|---|---|---|---|
| *NY-ESO-1* | CACACAGGATCCATGGATGCTGCAGATGCGG | 60 | 353 | 17 |
| | CACACAAAGCTTGGCTTAGCGCCTCTGCCCTG | | | 18 |
| *SP-10* | CCAGAGGAACATCAAGTCAGC | 64 | 964 | 19 |
| | GAGAAAGAGTTGGAGCAGGGAA | | | 20 |
| *SP17* | GGCAGTTCTTACCAAGAAGAT | 60 | 494 | 21 |
| | GGAGGTAAAACCAGTGTCCTC | | | 22 |
| *Acrosin* | TGCATGACTGGAGACTGGTT | 60 | 565 | 23 |
| | CAGTTCAGATAAGGCCAGGT | | | 24 |
| *PH-20* | AGAGGCCACTGAGAAAGCAA | 60 | 574 | 25 |
| | GGCTGCTAGTGTGACGTTGA | | | 26 |
| *OY-TES-1/sp32* | AAGGACAGGGGACTAAGGAG | 62 | 604 | 27 |
| | CCGTACAAATCCAGCCCGTA | | | 28 |
| *AKAP110* | CTAACTTCGGCCTTCCCAGA | 60 | 461 | 29 |
| | AGTGGGGTTGCCGATTACAG | | | 30 |
| *ASP* | AAGCAATTCACCAAGGCTGC | 60 | 552 | 31 |
| | ACCTATCATGCCGTTCTTCC | | | 32 |
| *Ropporin* | AGGTTCTACTGCTCTCCTTC | 60 | 631 | 33 |
| | GTAGAGAAACTGGAAGGTGC | | | 34 |
| *CS-1* | ATGGGAATGTGTGGCAGTAGA | 60 | 581 | 35 |
| | CCACTTACAATTTCCCGTCTG | | | 36 |
| *SPAG9* | ACTCCCACCAAAGGCATAGA | 60 | 515 | 37 |
| | CGAATCATCTCTGTCCATCG | | | 38 |
| *NYD-sp10* | TGTGTGACTCCATCCTCTAC | 60 | 640 | 39 |
| | AGGTAGAGCACGTAGTCATC | | | 40 |

[1] Forward primer sequence is shown in top and reverse primer sequence in bottom for each gene. Sequence is 5' – 3' for both primers.

To determine the specificity of these sperm-specific gene products as CT antigens, the expression of the corresponding genes in normal tissues was determined by RT-PCR of a panel of normal tissues. RT-PCR was conducted as described above.

**Results**

***Sperm protein mRNA expression in normal tissues by conventional RT-PCR.***

We investigated expression of sperm protein genes in normal tissues by RT-PCR analysis at 30 cycles. Eleven sperm protein genes (see Table 2) and well-defined control *NY-ESO-1* were amplified with 16 normal tissue cDNA templates (Multiple Tissue cDNA panel, CLONTECH). PCR products were analyzed by agarose gel electrophoresis and capillary electrophoresis on a microtip device by Agilent 2100 Bioanalyzer. As shown in Table 4, *acrosin*, *PH-20*, *OY-TES-1*, AKAP*110* and *NYD-sp10* mRNAs were amplified only in testis. *SP-10* and *ropporin* mRNA were amplified in testis and, to a lesser extent, in pancreas. *SP17*, *CS-1* and *SPAG9* mRNAs were amplified in most tissues.

***Real-time RT-PCR analysis of sperm protein genes in normal tissues***

To further analyze sperm protein mRNA expression in normal tissues, real-time RT-PCR analysis was performed. As shown in Fig. 1, CS-1 and SPAG9 showed mRNA expression in normal tissues ubiquitously, whereas other genes showed variable expression. Among tissues, the highest expression was consistently observed in testis. The gene with the highest expression in testis was *SP17*. Its threshold cycle (Ct ) value (*i.e.* the cycle at which the fluorescence of the reaction first arises above the background) was 21.8 for testis. Ct values of SP17 for other tissues, except skeletal muscle, were also rather high (26.9 - 30.4) (Fig. 1). The results were consistent with the above results obtained by conventional RT-PCR analysis.

The relative mRNA expression (n value, as described above) was determined. As shown in Fig. 2, *NY-ESO-1*, *SP-10*, *SP17*, *acrosin*, *PH-20*, *OY-TES-1*, *AKAP110*, *ASP*, *ropporin*, and *NYD-sp10* mRNA expression was $10^2$ to $10^7$ fold higher in testis to than in other tissues. *CS-1* mRNA was expressed 1.37, 1.63, and 8.13 fold higher in liver, placenta and pancreas, respectively, to that in testis. *SPAG9* mRNA expression in various tissues was 0.6-27% of that found in the testis.

Table 4: mRNA expression of sperm proteins in normal human tissues

| Tissues | Genes | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | OY-TES-1 (sp32) | SP-10 | SP17 | Acrosin | PH-20 | AKAP110 | ASP | Ropporin | CS-1 | SPAG9 | NYD-sp10 |
| Brain | - | - | + | - | - | - | - | - | + | + | - |
| Heart | - | - | + | - | - | - | - | - | + | ± | - |
| Kidney | - | - | + | - | - | - | - | - | - | - | - |
| Liver | - | - | + | - | - | - | - | - | + | + | - |
| Lung | - | - | + | - | - | - | - | - | + | ± | - |
| Pancreas | - | + | + | - | - | - | ± | + | + | ± | - |
| Placenta | - | - | + | - | - | - | - | - | + | - | - |
| Skeletal Muscle | - | - | + | - | - | - | - | - | + | - | - |
| Colon | - | - | + | - | - | - | - | - | + | + | - |
| Ovary | - | - | + | - | - | - | - | - | + | - | - |
| PBL | - | - | - | - | - | - | + | - | + | + | - |
| Prostate | - | - | + | - | - | - | - | - | + | - | - |
| Small Intestine | - | - | + | - | - | - | - | - | + | - | - |
| Spleen | - | - | + | - | - | - | ± | - | + | ± | - |
| Testis | + | + | + | + | + | + | + | + | + | + | + |
| Thymus | - | - | + | - | - | - | - | - | - | - | - |

***mRNA expression of selected sperm proteins in tumors***

Because of highly restricted mRNA expression in normal tissues, acrosin, PH-20, OY-TES-1, AKAP110, NYD-sp10, SP-10, and ropporin were chosen for mRNA expression analysis in malignant tissues by RT-PCR. The expression of the foregoing gene products was determined by RT-PCR of a panel of human tumor tissues. Samples of nine different types of cancer (bladder, breast, liver, lung, colon, stomach, renal, ovarian and glioma) were tested. As shown in Table 5, *AKAP110* mRNA was most frequently expressed in a variety of tumors. It was expressed in 26% (6/23) of bladder cancer samples, 20% (1/5) of liver cancer samples, 27% (4/15) of colon cancer samples, 40% (4/10) of renal cancer samples, and 39% (7/18) of ovarian cancer samples. No expression was observed in breast or stomach cancer samples. *Acrosin* was expressed in 5% (1/22) of bladder cancer samples, 20% (1/5) of breast cancer samples, 40% (2/5) of liver cancer samples, and 20% (1/5) of lung cancer samples. No expression of *acrosin* mRNA was observed in colon, stomach, renal and ovarian cancer samples. *SP-10, ropporin, PH-20* and *NYD-sp10* showed infrequent expression patterns in tumors.

These results indicated that five of the sperm proteins were specifically expressed in testis only: PH-20 (e.g., GenBank accession number XM_004865; SEQ ID NO:1, 2), AKAP110 (e.g., GenBank accession number AF093408; SEQ ID NO:3, 4), acrosin (e.g., GenBank accession number XM_010064; SEQ ID NO:5, 6), NYD-sp10 (e.g., GenBank accession number AF332192; SEQ ID NO:7, 8) and OY-TES-1 (previously determined to be a CT antigen (Ono et al., *Proc. Nat'l. Acad. Sci. USA* 98:3282-3287, 2001); e.g., GenBank accession number AB051833 (SEQ ID NO:41,42). In addition, two proteins, SP10 (e.g., GenBank accession number M82968 (SEQ ID NO:43,44) and ropporin (e.g., GenBank accession number NM_017578 (SEQ ID NO:45,46), were expressed in only testis and pancreas.

According to the expression pattern in normal and cancer tissues, the sperm-specific gene products PH-20, AKAP110, acrosin and NYD-sp10 were classified as additional CT antigens.

Table 5: mRNA expression of sperm specific proteins in human cancer

| Tumor type | Genes | | | | | | |
|---|---|---|---|---|---|---|---|
| | SP-10 | Acrosin | PH-20 | OY-TES-1/sp32 | AKAP110 | Ropporin | NYD-sp10 |
| Bladder cancer | 0/28 (0%) | 1/22 (5%) | 0/23 (0%) | 11/39 (28%) | 6/23 (26%) | N.D. | 0/22 (0%) |
| Breast cancer | 0/5 (0%) | 1/5 (20%) | 0/5 (0%) | 2/5 (40%) | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) |
| Liver cancer | 0/5 (0%) | 2/5 (40%) | 0/5 (0%) | 2/5 (40%) | 1/5 (20%) | 0/5 (0%) | 0/4 (0%) |
| Lung cancer | 1/5 (20%) | 1/5 (20%) | 0/5 (0%) | 1/5 (20%) | N.D. | 2/5 (40%) | 1/5 (20%) |
| Colon cancer | 0/15 (0%) | 0/15 (0%) | 0/15 (0%) | 2/13 (15%) | 4/15 (27%) | 0/15 (0%) | 0/15 (0%) |
| Stomach cancer | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) | 0/5 (0%) |
| Renal cancer | 0/10 (0%) | 0/10 (0%) | 0/10 (0%) | 0/10 (0%) | 4/10 (40%) | 0/10 (0%) | 0/10 (0%) |
| Ovarian cancer | 0/18 (0%) | 0/18 (0%) | 3/18 (17%) | 4/18 (22%) | 7/18 (39%) | 0/18 (0%) | 1/18 (6%) |
| Glioma | 7/34 (21%) | N.D. | 1/34 (3%) | 19/34 (56%) | 16/34 (47%) | 1/34 (3%) | 21/37 (57%) |

**Example 3: Expression of *RFX4* alternatively spliced variants in gliomas as cancer/testis antigens**

**Materials and Methods**

***Tissues***

Tumor tissues were obtained from patients who visited at Okayama University Medical School Hospital. Tumor specimens investigated in this study are listed in Table 6. For histological diagnosis of brain tumor specimens, World Health Organization (WHO) classification was used.

Table 6. *RFX4* mRNA expression in glioma and other tumors

| Tumor type | mRNA, positive/total |
|---|---|
| Glioblastoma | 21/37 (57%) |
| Astrocytoma G II | 3/9 (33%) |
| Astrocytoma G III | 8/11 (73%) |
| Astrocytoma G IV | 7/12 (58%) |
| Mixed glioma | 1/2 (50%) |
| Ependymoma | 2/3 (67%) |
| Meningioma | 0/8 (0%) |
| Lung cancer | 1/5 (20%) |
| Ovarian cancer | 1/20 (5%) |
| Cervical cancer | 1/16 (6%) |
| Breast cancer | 0/5 (0%) |
| Renal cancer | 0/10 (0%) |
| Bladder cancer | 0/22 (0%) |
| Liver cancer | 0/4 (0%) |
| Colon cancer | 0/15 (0%) |
| Stomach cancer | 0/5 (0%) |

**mRNA isolation and cDNA synthesis**

mRNA from frozen tumor tissues was purified using the QuickPrep Micro mRNA Purification Kit (Amersham Pharmacia, Piscataway, NJ). mRNA was reverse transcribed into single strand cDNA using Moloney murine leukemia virus reverse transcriptase and oligo $(dT)_{15}$ as a primer (Amersham Pharmacia). cDNAs were tested for integrity by amplification of β-actin transcripts in a 30 cycle reaction.

***Reverse-transcription PCR (RT-PCR)***

To amplify cDNA segments from normal tissues (Multiple Tissue cDNA panels, CLONTECH, Palo Alto, CA) and tumors, the gene specific primers listed in Table 7 were used. RT-PCR was performed by using 30 amplification cycles and followed by a 10-mi n elongation step at 72°C. The PCR products were analyzed by using conventional agarose gel electrophoresis.

***Rapid amplification of cDNA ends (RACE)***

5' RACE was performed to identify the 5' end sequence of *RFX4-C* using the 5'RACE System for Rapid Amplification kit (Gibco BRL, Rockville, MD). Total RNA was isolated from *RFX4-C* positive glioma specimens using the RNeasy kit (Qiagen GmbH, Hilden, Germany) and used as a template. The first-strand of cDNA was synthesized using the specific primer, GSP1-R1 (5'-CCCGAGTCTTCTGGTGGTTA-3') (SEQ ID NO:59). dC-tailed cDNA was amplified using a gene-specific nested primer GSP2-R1 (5'-AGCATTGACAGGTTGGGTATC-3') (SEQ ID NO:60) and an abridged universal anchor primer (5'-GGCCACGCGTCGACTAGTAC-3') (SEQ ID NO:61). The RACE product was sequenced with the sequence primer, RS1 (5'-AGTTCTCCTCCAGCCAT-3') (SEQ ID NO:62).

Table 7. Primer pairs used in this study

| Primer pairs | | Sequence of primers | Annealing temperature (°C) | PCR product size (bp) | SEQ ID NO: |
|---|---|---|---|---|---|
| A1 | A1-S | GCAATGGCTGGAGGAGAACT | 62 | 706 | 47 |
| | A1-AS | AGCCACTTTTAGCCACTTCATC | | | 48 |
| A2 | NYD-S | TGTGTGACTCCATCCTCTAC | 62 | 984 | 49 |
| | A2-AS | GTCTGGCTTTTTGTGTGTGTG | | | 50 |
| B1 | B1-S | GAAGACACGGAAGGCACAGA | 62 | 682 | 51 |
| | A1-AS | AGCCACTTTTAGCCACTCATC | | | 52 |
| B2 | B2-S | ACCGGAAACTCATCACCCCAAT | 62 | 1055 | 53 |
| | B2-AS | GTAAGCAAAGCCAGGAAAGTG | | | 54 |
| C1 | A1-S | GCAATGGCTGGAGGAGAACT | 62 | 1590 | 55 |
| | C1-AS | TAAACTGGTATCCTGTGTGTGA | | | 56 |
| common | NYD-S | TGTGTGACTCCATCCTCTAC | 60 | 640 | 57 |
| | NYD-AS | AGGTAGAGCACGTAGTCATC | | | 58 |

Forward primer sequence is shown in top and reverse primer sequence in bottom for each primer pair. Sequence is 5'-3' for both primers.

**Results**

***Expression of* RFX4 *mRNA in normal and malignant tissues***

*RFX4* gene is located on chromosome 12q24 and spans ~164-kb composed of 19 exons according to the NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map) (Fig. 3). Two alternatively spliced variants have been described. *RFX4-A* (SEQ ID NO:9, 10) that was originally described as *RFX4* by Morotomi-Yano et al. (51) and designated here as such is composed of exons 1-5, and 7-16, containing a DNA binding domain (DBD) encoded by exons 3, 4, 5 and 7 (Fig. 3 and 4). *RFX4-B*, which was reported as *NYD-sp10* (SEQ ID NO:7, 8) (GenBank accession number AF332192), is composed of exons 6-19 lacking DBD. Both products share evolutionarily conserved B, C regions and dimerization domain.

We investigated *RFX4* mRNA expression in adult normal tissues (Multiple Tissue cDNA panels, CLONTECH) and various tumors by RT-PCR using common primers for *RFX4-A* and *RFX4-B* (primer pair NYD-S and NYD-AS). As shown in Fig. 5, no expression of *RFX4* mRNA was observed in adult normal tissues except for testis. On the other hand, in tumors, a high level of *RFX4* mRNA expression was observed in gliomas. *RFX4* mRNA was detected in 33% (3/9) of astrocytoma G II, 73% (8/11) of astrocytoma G III, 58% (7/12) of astrocytoma G IV, 50% (1/2) of mixed glioma, and 67% (2/3) of ependymoma (Fig. 5 and Table 6). No expression was observed in meningiomas. In other tumors, *RFX4* mRNA was detected in 20% (1/5) of lung cancer, 5% (1/20) of ovarian cancer, and 6% (1/16) of cervical cancer. No expression of *RFX4* mRNA was observed in breast, renal, bladder, liver, colon, and stomach cancer.

***Expression of* RFX4 *alternatively spliced variants in glioma***

We further investigated the expression of alternatively spliced variants *RFX4-A* and *B* in gliomas using primer pairs as shown in Fig. 3 and Table 7. With 5' primer pairs A1 and B1, amplification was observed only with A1 in all 21 specimens of 37 gliomas that were positive for *RFX4* using common primers. However, with 3' primer pairs A2 and B2, amplification was observed by B2 only in the same 21 specimens. Amplification by primer pair A2 was observed in three tumor specimens. These results suggested that there is another splice variant in gliomas, designated *RFX4-C* (SEQ ID NOs:63 and 64 represent the

nucleotide and amino acid sequences, respectively), spanning the 5' end of *RFX4-A* to the 3' end of *RFX4-B* (Fig. 3).

We examined the expression of *RFX4-C* in gliomas using the *RFX4-C* specific primer pair C1 shown in Fig. 3. As shown in Fig. 6 and Table 8, all glioma specimens that were positive for *RFX4* using common primers also expressed *RFX4-C*. Expression of the splicing variants in various tumor specimens is shown in Table 8 below. 27% (3/8) of *RFX4-C* mRNA positive astrocytoma G III expressed *RFX4-A* simultaneously. No expression of *RFX4-B* was observed.

In testis, expression of *RFX4-A, B,* and *C* mRNA was observed.

Table 8. Expression of *RFX4* splicing variants in glioma

| Diagnosis | RFX4 positive specimens | RFX4-A | RFX4-B | RFX4-C |
|---|---|---|---|---|
| Astrocytoma G II | 3 | 0 | 0 | 3 |
| Astrocytoma G III | 8 | 3 | 0 | 8 |
| Astrocytoma G IV | 7 | 0 | 0 | 7 |
| Mixed glioma | 1 | 0 | 0 | 1 |
| Ependymoma | 2 | 0 | 0 | 2 |
| Total | 21 | 3 (14%) | 0 (0%) | 21 (100%) |

RT-PCR analysis was performed using primer pairs A1, A2, B1, B2 and C1 (Fig. 3 and Table 7) as shown in Fig. 6. All glioma specimens that were positive for *RFX4* using common primers in RT-PCR were also positive for *RFX4-C*. Three astrocytoma G III specimens expressed both *RFX4-A* and *C*.

**Example 4: Preparation of recombinant CT antigens**

To facilitate screening of patients' sera for antibodies or T cells reactive with CT antigens, for example by ELISA, recombinant proteins are prepared according to standard procedures. In one method, the clones encoding CT antigens are subcloned into a baculovirus expression vector, and the recombinant expression vectors are introduced into appropriate insect cells. Baculovirus/insect cloning systems are preferred because post-translational modifications are carried out in the insect cells. Another preferred eukaryotic system is the *Drosophila* Expression System from Invitrogen. Clones which express high

amounts of the recombinant protein are selected and used to produce the recombinant proteins. The recombinant proteins are tested for antibody recognition using serum from the patient which was used to isolated the particular clone, or in the case of CT antigens recognized by allogeneic sera, by the sera from any of the patients used to isolate the clones or sera which recognize the clones' gene products.

Alternatively, the CT antigen clones are inserted into a prokaryotic expression vector for production of recombinant proteins in bacteria. Other systems, including yeast expression systems and mammalian cell culture systems also can be used.

**Example 5: Preparation of antibodies to CT antigens**

The recombinant CT antigens produced as in Example 3 above are used to generate polyclonal antisera and monoclonal antibodies according to standard procedures. The antisera and antibodies so produced are tested for correct recognition of the CT antigens by using the antisera/antibodies in assays of cell extracts of patients known to express the particular CT antigen (e.g. an ELISA assay). These antibodies can be used for experimental purposes (e.g. localization of the CT antigens, immunoprecipitations, Western blots, etc.) as well as diagnostic purposes (e.g., testing extracts of tissue biopsies, testing for the presence of CT antigens).

The antibodies are useful for accurate and simple typing of cancer tissue samples for expression of the CT antigens.

**Example 6: Expression of CT antigens in cancers of similar and different origin.**

The expression of one or more of the CT antigens is tested in a range of tumor samples to determine which, if any, other malignancies should be diagnosed and/or treated by the methods described herein. Tumor cell lines and tumor samples are tested for CT antigen expression, preferably by RT-PCR according to standard procedures. Northern blots also are used to test the expression of the CT antigens. Antibody based assays, such as ELISA and western blot, also can be used to determine protein expression. A preferred method of testing expression of CT antigens (in other cancers and in additional same type cancer patients) is allogeneic serotyping using a modified SEREX protocol (as described above).

In all of the foregoing, extracts from the tumors of patients who provided sera for the initial isolation of the CT antigens are used as positive controls. The cells containing

recombinant expression vectors described in the Examples above also can be used as positive controls.

The results generated from the foregoing experiments provide panels of multiple cancer associated nucleic acids and/or polypeptides for use in diagnostic (e.g. determining the existence of cancer, determining the prognosis of a patient undergoing therapy, etc.) and therapeutic methods (e.g., vaccine composition, etc.).

**Example 7: HLA typing of patients positive for CT antigens**

To determine which HLA molecules present peptides derived from the CT antigens of the invention, cells of the patients which express the CT antigens are HLA typed. Peripheral blood lymphocytes are taken from the patient and typed for HLA class I or class II, as well as for the particular subtype of class I or class II. Tumor biopsy samples also can be used for typing. HLA typing can be carried out by any of the standard methods in the art of clinical immunology, such as by recognition by specific monoclonal antibodies, or by HLA allele-specific PCR (e.g. as described in WO97/31126).

**Example 8: Characterization of CT antigen peptides presented by MHC class I and class II molecules.**

Antigens which provoke an antibody response in a subject may also provoke a cell-mediated immune response. Cells process proteins into peptides for presentation on MHC class I or class II molecules on the cell surface for immune surveillance. Peptides presented by certain MHC/HLA molecules generally conform to motifs. These motifs are known in some cases, and can be used to screen the CT antigens for the presence of potential class I and/or class II peptides. Summaries of class I and class II motifs have been published (e.g., Rammensee et al., *Immunogenetics* 41:178-228, 1995). Based on the results of experiments such as those described above, the HLA types which present the individual CT antigens are known. Motifs of peptides presented by these HLA molecules thus are preferentially searched.

One also can search for class I and class II motifs using computer algorithms. For example, computer programs for predicting potential CTL epitopes based on known class I motifs has been described (*see, e.g.*, Parker et al, *J. Immunol.* 152:163, 1994; D'Amaro et al., *Human Immunol.* 43:13-18, 1995; Drijfhout et al., *Human Immunol.* 43:1-12, 1995). Computer programs for predicting potential T cell epitopes based on known class II motifs

has also been described (*see, e.g* Sturniolo et al., *Nat Biotechnol* 17(6):555-61, 1999). HLA binding predictions can conveniently be made using an algorithm available via the Internet on the National Institutes of Health World Wide Web site at URL http://bimas.dcrt.nih.gov . See also the website of: SYFPEITHI: An Internet Database for MHC Ligands and Peptide Motifs (access via http://www.uni-tuebingen.de/uni/kxi/ or http://134.2.96.221/scripts/hlaserver.dll/EpPredict.htm. Methods for determining HLA class II peptides and making substitutions thereto are also known (e.g. Strominger and Wucherpfennig (PCT/US96/03182)).

**Example 9: Identification of the portion of a cancer associated polypeptide encoding an antigen**

To determine if the CT antigens identified and isolated as described above can provoke a cytolytic T lymphocyte response, the following method is performed. CTL clones are generated by stimulating the peripheral blood lymphocytes (PBLs) of a patient with autologous normal cells transfected with one of the clones encoding a CT antigen polypeptide or with irradiated PBLs loaded with synthetic peptides corresponding to the putative protein and matching the consensus for the appropriate HLA class I molecule (as described above) to localize an antigenic peptide within the CT antigen clone (*see*, e.g., Knuth et al., *Proc. Natl. Acad. Sci. USA* 81:3511-3515, 1984; van der Bruggen et al., *Eur. J. Immunol.* 24:3038-3043, 1994). These CTL clones are screened for specificity against COS cells transfected with the CT antigen clone and autologous HLA alleles as described by Brichard et al. (*Eur. J. Immunol.* 26:224-230, 1996). CTL recognition of a CT antigen is determined by measuring release of TNF from the cytolytic T lymphocyte or by $^{51}$Cr release assay (Herin et al., *Int. J. Cancer* 39:390-396, 1987). If a CTL clone specifically recognizes a transfected COS cell, then shorter fragments of the CT antigen clone transfected in that COS cell are tested to identify the region of the gene that encodes the peptide. Fragments of the CT antigen clone are prepared by exonuclease III digestion or other standard molecular biology methods. Synthetic peptides are prepared to confirm the exact sequence of the antigen.

Optionally, shorter fragments of CT antigen cDNAs are generated by PCR. Shorter fragments are used to provoke TNF release or $^{51}$Cr release as above.

Synthetic peptides corresponding to portions of the shortest fragment of the CT antigen clone which provokes TNF release are prepared. Progressively shorter peptides are

synthesized to determine the optimal CT antigen tumor rejection antigen peptides for a given HLA molecule.

A similar method is performed to determine if the CT antigen contains one or more HLA class II peptides recognized by T cells. One can search the sequence of the CT antigen polypeptides for HLA class II motifs as described above. In contrast to class I peptides, class II peptides are presented by a limited number of cell types. Thus for these experiments, dendritic cells or B cell clones which express HLA class II molecules preferably are used.

References

1. Boyse EA, Miyazawa M, Aoki T, Old LJ. Ly-A and Ly-B: two systems of lymphocyte isoantigens in the mouse. *Proc Royal Soc Brit* 1968; 170: 175-193. (PMID:4385242)

2. Boyse EA, Old LJ. Some aspects of normal and abnormal cell surface genetics. *Ann Rev Genet* 1969; 3: 269-290.

3. DeLeo AB, Jay G, Appella E, Dubois GC, Law LW, Old LJ. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc Natl Acad Sci USA* 1979; 76: 2420-2424. (PMID: 221923)

4. Lilly F, Boyse EA, Old LJ. Genetic basis of susceptibility to viral leukemogenesis. *Lancet* 1965; ii:1207-1209.

5. Carswell EA, Old LJ, Kassel RL, Green S, Fiore NC, Williamson B. An endotoxin-induced serum factor that causes necrosis of tumors. *Proc Natl Acad Sci USA* 1975; 72: 3666-3670. (PMID: 1103152)

6. Traversari C, van der Bruggen P, Van den Eynde B, Hainaut P, Lemoine C, Ohta N, Old LJ, Boon T. Transfection and expression of a gene coding for a human melanoma antigen recognized by autologous cytolytic T lymphocytes. *Immunogenetics* 1992; 35: 145-152. (PMID: 1537606)

7. van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, Knuth A, and Boon, T, A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* 1991; 254:1643-1647. (PMID: 1840703)

8. Old LJ. Cancer immunology: The search for specificity. *Cancer Res* 1981; 41: 361-375. (PMID: 7004632)

9. Knuth A, Danowski B, Oettgen HF, Old LJ. T cell-mediated cytotoxicity against malignant melanoma: Analysis with IL-2-dependent T cell cultures. *Proc Natl Acad Sci USA* 1984; 81: 3511-3515. (PMID: 6610177)

10. Sahin, U, Türeci Ö, Schmitt, Cochlovius B, Johannes T, Schmits R, Stenner F, Luo G, Schobert I, Pfreundschuh M. Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc Natl Acad Sci USA* 1995; 92:11810-11813. (PMID: 8524854)

11. Old LJ, Chen YT. New paths in human cancer serology. *J Exp Med* 1998; 187:1163-1167. (PMID: 9547328)

12. De Plaen E, Arden K, Traversari C, Gaforio JJ, Szikora JP, De Smet C, Brasseur F, van der Bruggen P, Lethé B, Lurquin C, Brasseur R, Chomez P, De Backer O, Cavenee W, and Boon T. Structure, chromosomal localization and expression of twelve genes of the MAGE family. *Immunogenetics* 1994; 40: 360-369. (PMID: 7927540)

13. Muscatelli F, Walker AP, De Plaen E, Stafford AN, Monaco AP. Isolation and characterization of a new MAGE gene family in the Xp21.3 region. *Proc Natl Acad Sci USA* 1995; 92: 4987-4991. (PMID: 7761436)

14. Boël P, Wildmann C, Sensi ML, Brasseur R, Renauld JC, Coulie P, Boon T, van der Bruggen P. BAGE, a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes. *Immunity* 1995; 2: 167-175. (PMID: 7895173)

15. Van den Eynde B, Peeters O, De Backer O, Gaugler B, Lucas S, Boon T. A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma. *J Exp Med* 1995;182: 689-698. (PMID: 7544395)

16. De Backer O, Arden KC , Boretti M, Vantomee V, De Smet C, Czekay S, Viars CS, De Plaen E, Brasseur F, Chomez P, Van den Eynde B, Boon T, van der Bruggen P. Characterization of the GAGE genes that are expressed in various human cancer and in normal testis. *Cancer Res* 1999; 59: 3157-65. (PMID: 10397259)

17. Güre AO, Türeci Ö, Sahin U, Tsang S, Scanlan MJ, Jager E, Knuth A, Pfreundschuh M, Old LJ, Chen YT. SSX, a multigene family with several members transcribed in normal testis and human cancer. *Int. J. Cancer* 1997; 72: 965-971. (PMID: 9378559)

18. Chen YT, Scanlan MJ, Sahin U, Türeci Ö, Güre AO, Tsang S, Williamson B, Stockert E, Pfreundschuh M, Old LJ. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc. Natl. Acad. Sci USA* 1997; 94: 1914-1918. (PMID: 9050879)

19. Lethe B, Lucas S, Michaux L, De Smet C, Godelaine D, Serrano A, De Plaen E, Boon T. LAGE-1: a new gene with tumor specificity. *Int J Cancer* 1998;76:903-908. (PMID: 9626360)

20. Türeci Ö, Dahin U, Zwick C, Koslowski M, Switz G, Pfreundschuh M. Identification of a meiosis-specific protein as a new member of the class of cancer/testis antigens. *Proc Natl Acad Sci USA* 1998; 95: 5211-5216. (PMID: 9560255)

21. Chen YT, Güre AO, Tsang S, Stockert E, Jäger E, Knuth A, Old LJ. Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library. *Proc Natl Acad Sci USA* 1998; 95: 6919-6923. (PMID: 9618514)

22. Lucas S, De Smet C, Arden KC, Viars CS, Lethe B, Lurquin C, Boon T. Identification

of a new MAGE gene with tumor-specific expression by representational difference analysis. *Cancer Res* 1998; 58: 743-752. (PMID: 9485030)

23. Sahin U, Koslowski M., Türeci Ö, Eberle T, Zwick C, Romeike B, Moringlane JR, Schwechheimer K, Feiden W., Pfreundschuh M. Expression of cancer/testis genes in human brain tumors. *Clin Cancer Res* 2000;10: 3916-3922. (PMID: 11051238)

24. Scanlan MJ, Altorki NK, Güre AO, Williamson B, Jungbluth A, Chen YT, Old LJ. Expression of cancer-testis antigens in lung cancer: definition of bromodomain testis-specific gene (BRDT) as a new CT gene CT9. *Cancer Lett.* 2000; 150:155-164. (PMID: 10704737)

25. Güre AO, Stockert E, Arden KC, Boyer AD, Viars CS, Scanlan MJ, Old LJ, Chen YT. CT10: a new cancer-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational difference analysis. *Int J Cancer* 2000; 85: 726-732. (PMID: 10699956)

26. Lucas S, De Plaen E, Boon T. MAGE-B5, MAGE-B6, MAGE-C2 and MAGE-C3: four new member of the MAGE family with tumor-specific expression. *Int J Cancer* 2000; 87 :55-60. (PMID: 10861452)

27. Zendman AJ, Cornelissen IM, Weidle UH, Ruiter DJ, van Muijen GN. CTp11, a novel member of the family of human cancer/testis antigens. *Cancer Res* 1999, 59: 6223-6239. (PMID: 10626816)

28. Martelange V, De Smet C, De Palen E, Lurquin C, Boon, T. Identification on a human sarcoma of two new genes with tumor-specific expression. *Cancer Res* 2000; 60: 3848-3855. (PMID: 10919659)

29. Eichmüller S, Usener D, Dummer R, Stein A, Thiel D, Schadendorf D. Serological detection of cutaneous T cell lymphoma-associated antigens. *Proc Natl Acad Sci USA* 2001; 98: 629-634. (PMID: 11149944)

30. Ono T, Kurashige T, Harada N, Noguchi Y, Saika T, Niikawa N, Aoe M, Nakamura S, Higashi T, Hiraki A, Wada H, Kumon H, Old L, Nakayama E. Identification of proacrosin binding protein sp32 precursor as a human cancer/testis antigen. *Proc Natl Acad Sci.USA* 2001; 98:3 282-3287. (PMID: 11248070)

31. Jungbluth A, Busam K, Kolb D, Iversen K, Coplan K, Chen YT, Spagnoli GC, Old LJ. Expression of MAGE-antigens in normal tissues and cancer. *Int J Cancer*. 2000; 85 :460-5. ( PMID: 10699915)

32. Jungbluth A, Busam K, Iversen, K, Kolb D, Coplan K, Chen YT, Stockert E, Zhang P, Old, LJ. Cancer-Testis (CT) antigens MAGE-1, MAGE-3, NY-ESO-1, and CT7 are expressed in female germ cells. *Mod Path*. In press 2001.

33. Jungbluth A, Iversen K, Kolb D, Coplan K, Chen YT, Stockert E, Old LJ, Vogel M. Expression of CT (Cancer/Testis) antigens MAGE, NY-ESO-1, and CT7 in placenta. *German Soc for Path*. Submitted 2001.

34. Jungbluth A, Stockert E, Chen YT, Kolb D, Iversen K, Coplan K, Williamson B, Altorki N, Busam KJ, Old LJ. Monoclonal antibody MA454 reveals a heterogeneous expression pattern of MAGE-1 antigen in formalin-fixed paraffin embedded lung tumours. *Br J Cancer* 2000; 83: 493-497. (PMID: 10945497)

35. Meuwissen RJL, Offenberg, HH, Dietrich AJ, Riesewijk A, van Iersel M, Heting C. A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. *EMBO J* 1992;11: 5091-5100. (PMID: 1464329)

36. Baba T, Niida Y, Michikawa Y, Kashiwabara S, Kodaira K, Takenaka M, Kohno N, Gerton GL, Arai Y. An acrosomal protein, sp32, in mammalian sperm is a binding protein specific for two proacrosins and an acrosin intermediate. *J Biochem* 1994; 269:10133-10140. (PMID: 8144514)

37. Brasseur F, Rimoldi D, Liénard D, Lethe B, Carrel S, Arienti F, Suter L, Vanwijck R, Bourlond A, Humblet Y, Vacca A, Conese M, Lahaye T, Degiovanni G,

Deraemaecker R, Beauduin M, Sastre X, Salamon E, Dréno B, Jäger E, Knuth A, Chevreau C, Suciu S, Lachapelle J-M, Pouillart P, Parmiani G, Lejeune F, Cerottini J-C, Boon T, Marchand M. Expression of MAGE gene in primary and metastatic cutaneous melanoma. *Int J Cancer* 1995; 63:375-380. (PMID: 7591235)

38. Patard JJ, Brasseur F, Gil-Diez S, Radvanyi F, Marchand M, Francois P, Abi-Aad A, VanCangh P, Abbou CC, Chopin D, Boon T. Expression of MAGE genes in transitional-cell carcinomas of the urinary bladder. *Int J Cancer* 1995; 64:60-64. PMID: 7665250)

39. Kurashige T, Noguchi Y, Saika T, Ono T, Nagata Y, Jungluth A, Ritter G, Chen YT, Stockert E, Tomoyasu T, Kumon H, Old LJ, Nakayama E. NY-ESO-1 expression and immunogenicity associated with transitional cell carcinoma: correlation with tumor grade. *Cancer Res* 61:4671-4674, 2001.

40. Stockert E., Jäger E, Chen YT, Scanlan M, Gout I, Karbach J, Arand M, Knuth A, Old, LJ. A survey of the humoral response of cancer patients to a panel of human tumor antigens. *J Exp Med* 1998;187:1349-1354. (PMID: 9547346)

41. Jäger E., Nagata Y, Gnjatic S, Wada H, Stockert E, Karbach J, Dunbar PR, Lee SY, Jungbluth A, Jäger D, Arand M, Ritter G, Cerundolo V, Dupont B, Chen YT, Old LJ, Knuth A. Monitoring CD8 T cell responses to NY-ESO-1: Correlation of humoral and cellular immune responses. *Proc Natl Acad Sci USA* 2000; 97: 4760-4765. (PMID: 10781081)

42. De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with geneome-wide demethylation. *Proc. Natl. Acad. Sci USA* 1996; 93:7149-7153 (.PMID: 8692960)

43. Jungbluth A, Chen YT, Stockert E, Busam KJ, Kolb D, Iversen K, Coplan K, Williamson B, Altorki N, Old LJ. Immunohistochemical analysis of NY-ESO-1 antigen expression in normal and malignant tumors. *Int J Cancer.* 92:856-860, 2001.

44. Jungbluth A, Antonescu C, Busam K, Iversen K, Kolb D, Coplan K, Chen YT, Stockert E, Ladanyi M, Old, LJ. Monophasic and biphasic synovial sarcomas abundantly express cancer/testis antigen NY-ESO-1, but not MAGE-A1 or CT7. *Int J Cancer*. 94:252-256, 2001.

45. Antonescu C, Busam K, Iversen K, Kolb D, Coplan K, Spagnoli G, Ladanyi M, Old LJ, Jungbluth, A. MAGE antigen expression in monophasic and biphasic synovial sarcoma. *Mod Path.* Submitted 2001.

46. Beard J. The cancer problem. *Lancet* 1905;1:281-203.

47. Gurchot C. The trophoblast theory of cancer. *Oncology* 1975; 31: 310-333. (PMID: 1107920)

48. Iles RK, Chard T. Human Chorionic Gonadotropin Expression by Bladder Cancers: Biology and Clinical Potential. *J Urol* 1991; 145 :453-458. (PMID: 1705292)

49. Acevedo HF, Tong JY, Hartsock RJ. Human Chorionic Gonadotropin-Beta Subunit Gene Expression in Cultured Human Fetal and Cancer Cells of Different Types and Origins. *Cancer* 1995; 76: 1467-1475. (PMID: 8620425)

50. Dirnhofer S, Koessler P, Ensigner C, Feichtinger H, Madersbacher S, Berger P. Production of Trophoblastic Hormones by Transitional Cell Carcinoma of the Bladder: Association to Tumor Stage and Grade. *Hum Path* 1998; 29: 377-382. (PMID: 9563788)

51. Morotomi-Yano K, Yano K, Saito H, Sun Z, Iwama A, Miki Y. Human Regulatory Factor X 4 (RFX4) Is a Testis-specific Dimeric DNA-binding Protein That Cooperates with Other Human RFX Members. *J Biol Chem* 2002; 277(1): 836-842.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All references disclosed herein are incorporated by reference in their entirety.

We claim:

## Claims

1. A method of diagnosing a disorder characterized by expression of a human CT antigen precursor coded for by a nucleic acid molecule, comprising:

contacting a biological sample isolated from a subject with an agent that specifically binds to the nucleic acid molecule, an expression product thereof, a fragment of an expression product thereof complexed with an HLA molecule, or an antibody that binds the expression product thereof, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 and

determining the interaction between the agent and the nucleic acid molecule, the expression product or the antibody as a determination of the disorder.

2. The method of claim 1, wherein the agent is selected from the group consisting of

(a) a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 or a fragment thereof,

(b) an antibody that binds to an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63,

(c) an agent that binds to a complex of an HLA molecule and a fragment of an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and

(d) an expression product of a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 that binds an antibody.

3. The method of claim 1, wherein the disorder is characterized by expression of a plurality of human CT antigen precursors and wherein the agent is a plurality of agents, each of which is specific for a different human CT antigen precursor, and wherein said plurality of agents is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8, at least 9 or at least 10 such agents.

4. The method of claims 1-3, wherein the disorder is cancer.

5. The method of claim 1, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1.

6. The method of claim 1, wherein nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3.

7. A method for determining regression, progression or onset of a condition characterized by expression of abnormal levels of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, comprising

monitoring a sample, from a patient who has or is suspected of having the condition, for a parameter selected from the group consisting of

(i) the protein,

(ii) a peptide derived from the protein,

(iii) an antibody which selectively binds the protein or peptide, and

(iv) cytolytic T cells specific for a complex of the peptide derived from the protein and an MHC molecule,

as a determination of regression, progression or onset of said condition.

8. The method of claim 7, wherein the sample is a body fluid, a body effusion, cell or a tissue.

9. The method of claim 7, wherein the step of monitoring comprises contacting the sample with a detectable agent selected from the group consisting of

(a) an antibody which selectively binds the protein of (i), or the peptide of (ii),

(b) a protein or peptide which binds the antibody of (iii), and

(c) a cell which presents the complex of the peptide and MHC molecule of (iv).

10. The method of claim 9, wherein the antibody, the protein, the peptide or the cell is labeled with a radioactive label or an enzyme.

11. The method of claim 7, comprising assaying the sample for the peptide.

12. The method of claim 7, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1.

13. The method of claim 7, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3.

14. The method of claim 7, wherein the protein is a plurality of proteins, the parameter is a plurality of parameters, each of the plurality of parameters being specific for a different of the plurality of proteins, at least one of which is a CT antigen protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

15. The method of claim 7, wherein the protein is a plurality of proteins, at least one of which is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and wherein the parameter is a plurality of parameters, each of the plurality of parameters being specific for a different of the plurality of proteins.

16. A pharmaceutical preparation for a human subject comprising

an agent which when administered to the subject enriches selectively the presence of complexes of an HLA molecule and a human CT antigen peptide, and

a pharmaceutically acceptable carrier, wherein the human CT antigen peptide is a fragment of a human CT antigen encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

17. The pharmaceutical preparation of claim 16, wherein the agent comprises a plurality of agents, each of which enriches selectively in the subject complexes of an HLA molecule and a different human CT antigen peptide, wherein at least one of the human CT antigens is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

18. The pharmaceutical preparation of claim 17, wherein the plurality is at least two, at least three, at least four or at least five different such agents.

19. The pharmaceutical preparation of claim 16, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

20. The pharmaceutical preparation of claim 16, wherein the agent comprises a plurality of agents, at least one of which is a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3, or an expression product thereof, each of which enriches selectively in the subject complexes of an HLA molecule and a different human CT antigen.

21. The pharmaceutical preparation of claim 14, wherein the agent is selected from the group consisting of

(1) an isolated polypeptide comprising the human CT antigen peptide, or a functional variant thereof,

(2) an isolated nucleic acid operably linked to a promoter for expressing the isolated polypeptide, or functional variant thereof,

(3) a host cell expressing the isolated polypeptide, or functional variant thereof, and

(4) isolated complexes of the polypeptide, or functional variant thereof, and an HLA molecule.

22. The pharmaceutical preparation of claims 16-21, further comprising an adjuvant.

23. The pharmaceutical preparation of claim 16, wherein the agent is a cell expressing an isolated polypeptide comprising the human CT antigen peptide or a functional variant thereof, and wherein the cell is nonproliferative.

24. The pharmaceutical preparation of claim 16, wherein the agent is a cell expressing an isolated polypeptide comprising the human CT antigen peptide or a functional variant thereof, and wherein the cell expresses an HLA molecule that binds the polypeptide.

25. The pharmaceutical preparation of claim 23 or 24, wherein the isolated polypeptide comprises a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

26. The pharmaceutical preparation of claim 16, wherein the agent is at least two, at least three, at least four or at least five different polypeptides, each coding for a different human CT antigen peptide or functional variant thereof, wherein at least one of the human CT antigen peptidess is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

27. The pharmaceutical preparation of claim 26, wherein the at least one of the human CT antigen peptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3, or a fragment thereof.

28. The pharmaceutical preparation of claim 16, wherein the agent is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:1.

29. The pharmaceutical preparation of claim 16, wherein the agent is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:3.

30. The pharmaceutical preparation of claim 24, wherein the cell expresses one or both of the polypeptide and HLA molecule recombinantly.

31. The pharmaceutical preparation of claim 24, wherein the cell is nonproliferative.

32. A composition comprising

an isolated agent that binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

33. The composition of matter of claim 32, wherein the agent binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:1.

34. The composition of matter of claim 32, wherein the agent binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:3.

35. The composition of matter of claim 32, wherein the agent binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:5.

36. The composition of matter of claim 32, wherein the agent binds selectively a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence set forth as SEQ ID NO:7.

37. The composition of matter of claims 32-36, wherein the agent is a plurality of different agents that bind selectively at least two, at least three, at least four, or at least five different such polypeptides.

38. The composition of matter of claim 37, wherein the at least one of the polypeptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3, or a fragment thereof.

39. The composition of matter of claims 32-36, wherein the agent is an antibody.

40. The composition of matter of claim 37, wherein the agent is an antibody.

41. A composition of matter comprising
a conjugate of the agent of claims 32-36 and a therapeutic or diagnostic agent.

42. A composition of matter comprising
a conjugate of the agent of claim 37 and a therapeutic or diagnostic agent.

43. The composition of matter of claim 41, wherein the conjugate is of the agent and a therapeutic or diagnostic that is a toxin.

44. A pharmaceutical composition comprising an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and a pharmaceutically acceptable carrier.

45. The pharmaceutical composition of claim 44, wherein the isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

46. The pharmaceutical composition of claim 44, wherein the isolated nucleic acid molecule comprises at least two isolated nucleic acid molecules coding for two different polypeptides, each polypeptide comprising a different human CT antigen.

47. The pharmaceutical composition of claim 46, wherein at least one of the nucleic acid molecules comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs:1 and 3.

48. The pharmaceutical composition of claims 44-47 further comprising an expression vector with a promoter operably linked to the isolated nucleic acid molecule.

49. The pharmaceutical composition of claims 44-47 further comprising a host cell recombinantly expressing the isolated nucleic acid molecule.

50. A pharmaceutical composition comprising

an isolated polypeptide comprising a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, and

a pharmaceutically acceptable carrier.

51. The pharmaceutical composition of claim 50, wherein the isolated polypeptide comprises a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

52. The pharmaceutical composition of claim 50, wherein the isolated polypeptide comprises at least two different polypeptides, each comprising a different human CT antigen.

53. The pharmaceutical composition of claim 52, wherein at least one of the polypeptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:1 and 3.

54. The pharmaceutical composition of claims 50-53, further comprising an adjuvant.

55. A protein microarray comprising at least one polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or an antigenic fragment thereof.

56. The microarray of claim 55, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1.

57. The microarray of claim 55, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3.

58. The microarray of claim 55, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:5.

59. The microarray of claim 55, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:7.

61. A protein microarray comprising an antibody or an antigen-binding fragment thereof that specifically binds at least one polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or an antigenic fragment thereof.

62. The microarray of claim 61, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1.

63. The microarray of claim 61, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3.

64. The microarray of claim 61, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:5.

65. The microarray of claim 61, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:7.

67. A nucleic acid microarray comprising at least one nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample.

68. The microarray of claim 67, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample.

69. The microarray of claim 67, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample.

70. The microarray of claim 67, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:5, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample.

71. The microarray of claim 67, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:7, or a fragment thereof of at least 20 nucleotides that selectively hybridizes to its complement in a biological sample.

73. An isolated fragment of a human CT antigen which, or a portion of which, binds a HLA molecule or a human antibody, wherein the CT antigen is encoded by a nucleic acid

molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

74. The fragment of claim 73, wherein the fragment is part of a complex with the HLA molecule.

75. The fragment of claim 73, wherein the fragment is between 8 and 12 amino acids in length.

76. A kit for detecting the expression of a human CT antigen comprising

a pair of isolated nucleic acid molecules each of which consists essentially of a molecule selected from the group consisting of (a) a 12-32 nucleotide contiguous segment of the nucleotide sequence of any of SEQ ID NOS:1, 3, 5, 7, 9 and 63 and (b) complements of (a), wherein the contiguous segments are nonoverlapping.

77. The kit of claim 76, wherein the pair of isolated nucleic acid molecules is constructed and arranged to selectively amplify an isolated nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 and 3.

78. A method for treating a subject with a disorder characterized by expression of a human CT antigen, comprising

administering to the subject an amount of an agent, which enriches selectively in the subject the presence of complexes of a HLA molecule and a human CT antigen peptide, effective to ameliorate the disorder, wherein the human CT antigen peptide is a fragment of a human CT antigen encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

79. The method of claim 78, wherein the disorder is characterized by expression of a plurality of human CT antigens and wherein the agent is a plurality of agents, each of which enriches selectively in the subject the presence of complexes of an HLA molecule and a different human CT antigen peptide, wherein at least one of the human CT antigens is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

80. The method of claim 79, wherein at least one of the human CT antigen peptides is a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3, or a fragment thereof.

81. The method of claim 79, wherein the plurality is at least 2, at least 3, at least 4, or at least 5 such agents.

82. The method of claims 78-81, wherein the agent is an isolated polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

83. The method of claims 78-81, wherein the disorder is cancer.

84. The method of claims 82, wherein the disorder is cancer.

85. A method for treating a subject having a condition characterized by expression of a human CT antigen in cells of the subject, comprising:

(i) removing an immunoreactive cell containing sample from the subject,

(ii) contacting the immunoreactive cell containing sample to the host cell under conditions favoring production of cytolytic T cells against a human CT antigen peptide that is a fragment of the human CT antigen,

(iii) introducing the cytolytic T cells to the subject in an amount effective to lyse cells which express the human CT antigen, wherein the host cell is transformed or transfected with an expression vector comprising an isolated nucleic acid molecule operably linked to a promoter, wherein the isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

86. The method of claim 85, wherein the host cell recombinantly expresses an HLA molecule which binds the human CT antigen peptide.

87. The method of claim 85, wherein the host cell endogenously expresses an HLA molecule which binds the human CT antigen peptide.

88. A method for treating a subject having a condition characterized by expression of a human CT antigen in cells of the subject, comprising:

(i) identifying a nucleic acid molecule expressed by the cells associated with said condition, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63;

(ii) transfecting a host cell with a nucleic acid selected from the group consisting of (a) the nucleic acid molecule identified, (b) a fragment of the nucleic acid identified which includes a segment coding for a human CT antigen, (c) deletions, substitutions or additions to (a) or (b), and (d) degenerates of (a), (b), or (c);

(iii) culturing said transfected host cells to express the transfected nucleic acid molecule, and;

(iv) introducing an amount of said host cells or an extract thereof to the subject effective to increase an immune response against the cells of the subject associated with the condition.

89. The method of claim 88, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

90. The method of claim 88, further comprising identifying an MHC molecule which presents a portion of an expression product of the nucleic acid molecule, wherein the host cell expresses the same MHC molecule as identified and wherein the host cell presents an MHC binding portion of the expression product of the nucleic acid molecule.

91. The method of claim 88, wherein the immune response comprises a B-cell response or a T cell response.

92. The method of claim 91, wherein the response is a T-cell response which comprises generation of cytolytic T-cells specific for the host cells presenting the portion of the expression product of the nucleic acid molecule or cells of the subject expressing the human CT antigen.

93. The method of claim 88, wherein the nucleic acid molecule is selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63.

94. The method of claims 88 or 90, further comprising treating the host cells to render them non-proliferative.

95. A method for treating or diagnosing or monitoring a subject having a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta, comprising

administering to the subject an antibody which specifically binds to the protein or a peptide derived therefrom, the antibody being coupled to a therapeutically useful agent, in an amount effective to treat the condition.

96. The method of claim 95, wherein the antibody is a monoclonal antibody.

97. The method of claim 96, wherein the monoclonal antibody is a chimeric antibody or a humanized antibody.

98. A method for treating a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta, comprising

administering to a subject a pharmaceutical composition of any one of claims 16-31 and 44-54 in an amount effective to prevent, delay the onset of, or inhibit the condition in the subject.

99. The method of claim 98, wherein the condition is cancer.

100. The method of claim 98, further comprising first identifying that the subject expresses in a tissue abnormal amounts of the protein.

101. The method of claim 99, further comprising first identifying that the subject expresses in a tissue abnormal amounts of the protein.

102. A method for treating a subject having a condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta, comprising

(i) identifying cells from the subject which express abnormal amounts of the protein;

(ii) isolating a sample of the cells;

(iii) cultivating the cells, and

(iv) introducing the cells to the subject in an amount effective to provoke an immune response against the cells.

103. The method of claim 102, further comprising rendering the cells non-proliferative, prior to introducing them to the subject.

104. A method for treating a pathological cell condition characterized by expression of a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63 in cells or tissues other than testis, fetal ovary or placenta, comprising

administering to a subject in need thereof an effective amount of an agent which inhibits the expression or activity of the protein.

105. The method of claim 104, wherein the agent is an inhibiting antibody which selectively binds to the protein and wherein the antibody is a monoclonal antibody, a chimeric antibody, a humanized antibody or an antibody fragment.

106. The method of claim 104, wherein the agent is an antisense nucleic acid molecule which selectively binds to the nucleic acid molecule which encodes the protein.

107. The method of claim 104, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:1.

108. The method of claim 104, wherein the nucleic acid molecule comprises a nucleotide sequence set forth as SEQ ID NO:3.

109. A composition of matter useful in stimulating an immune response to a plurality of a proteins encoded by nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 63, comprising

a plurality of peptides derived from the amino acid sequences of the proteins, wherein the peptides bind to one or more MHC molecules presented on the surface of cells which are not testis, fetal ovary or placenta.

110. The composition of matter of claim 109, wherein at least a portion of the plurality of peptides bind to MHC molecules and elicit a cytolytic response thereto.

111. The composition of matter of claim 109, wherein at least one of the proteins is encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3.

112. The composition of matter of claim 110, further comprising an adjuvant.

113. The composition of matter of claim 112, wherein said adjuvant is a saponin, GM-CSF, or an interleukin.

114. The composition of matter of claim 109, further comprising at least one peptide useful in stimulating an immune response to at least one protein which is not encoded by SEQ ID NOS:1, 3, 5, 7, 9 and 63, wherein the at least one peptide binds to one or more MHC molecules.

115. An isolated antibody which selectively binds to a complex of:

(i) a peptide derived from a protein encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting SEQ ID NOS:1, 3, 5, 7, 9 and 63 and

(ii) and an MHC molecule to which binds the peptide to form the complex,

wherein the isolated antibody does not bind to (i) or (ii) alone.

116. The antibody of claim 115, wherein the antibody is a monoclonal antibody, a chimeric antibody, a humanized antibody, or a fragment thereof.

117. A method for identifying nucleic acids that encode a CT antigen, comprising

screening sequence database records for sequences that are expressed in a first set of samples consisting of cancers of at least two tissues and are expressed in a second set of samples consisting of at least one tissue selected from the group consisting of testis, ovary and placenta,

identifying as CT antigens the sequences that match the expression criteria.

118. The method of claim 117, wherein the sequences are expressed in cancers at least three tissues.

119. The method of claim 117, wherein the second tissue is testis.

120. The method of claim 117, wherein the second tissue is ovary.

121. The method of claim 120, wherein the second tissue is fetal ovary.

122. The method of claim 117, further comprising verifying the expression pattern of the sequences in normal tissue samples and/or tumor samples.

123. The method of claim 122, wherein the expression pattern is verified by nucleic acid amplification or nucleic acid hybridization.

124. A method for identifying nucleic acids that encode a CT antigen, comprising

screening sequence database records for sequences that are expressed in a first set of samples consisting of cancers of at least two tissues and are gamete-specific gene products,

identifying as CT antigens the sequences that match the expression criteria.

125. The method of claim 124, wherein the sequences are expressed in cancers at least three tissues.

126. The method of claim 124, further comprising verifying the expression pattern of the sequences in normal gamete tissue samples and/or tumor samples.

127. The method of claim 126, wherein the expression pattern is verified by nucleic acid amplification or nucleic acid hybridization.

128. A method for identifying nucleic acids that encode a CT antigen, comprising

screening sequence database records for sequences that are expressed in a first set of samples consisting of cancers of at least two tissues and are gene products associated with meiosis,

identifying as CT antigens the sequences that match the expression criteria.

129. The method of claim 128, wherein the sequences are expressed in cancers at least three tissues.

130. The method of claim 128, further comprising verifying the expression pattern of the sequences in normal meiotic tissue samples and/or tumor samples.

131. The method of claim 130, wherein the expression pattern is verified by nucleic acid amplification or nucleic acid hybridization.

132. A method for identifying nucleic acids that encode a CT antigen, comprising

screening sequence database records for sequences that are expressed in a first set of samples consisting of cancers of at least two tissues and are trophoblast-specific gene products,

identifying as CT antigens the sequences that match the expression criteria.

133. The method of claim 132, wherein the sequences are expressed in cancers at least three tissues.

134. The method of claim 132, further comprising verifying the expression pattern of the sequences in normal trophoblast tissue samples and/or tumor samples.

135. The method of claim 134, wherein the expression pattern is verified by nucleic acid amplification or nucleic acid hybridization.

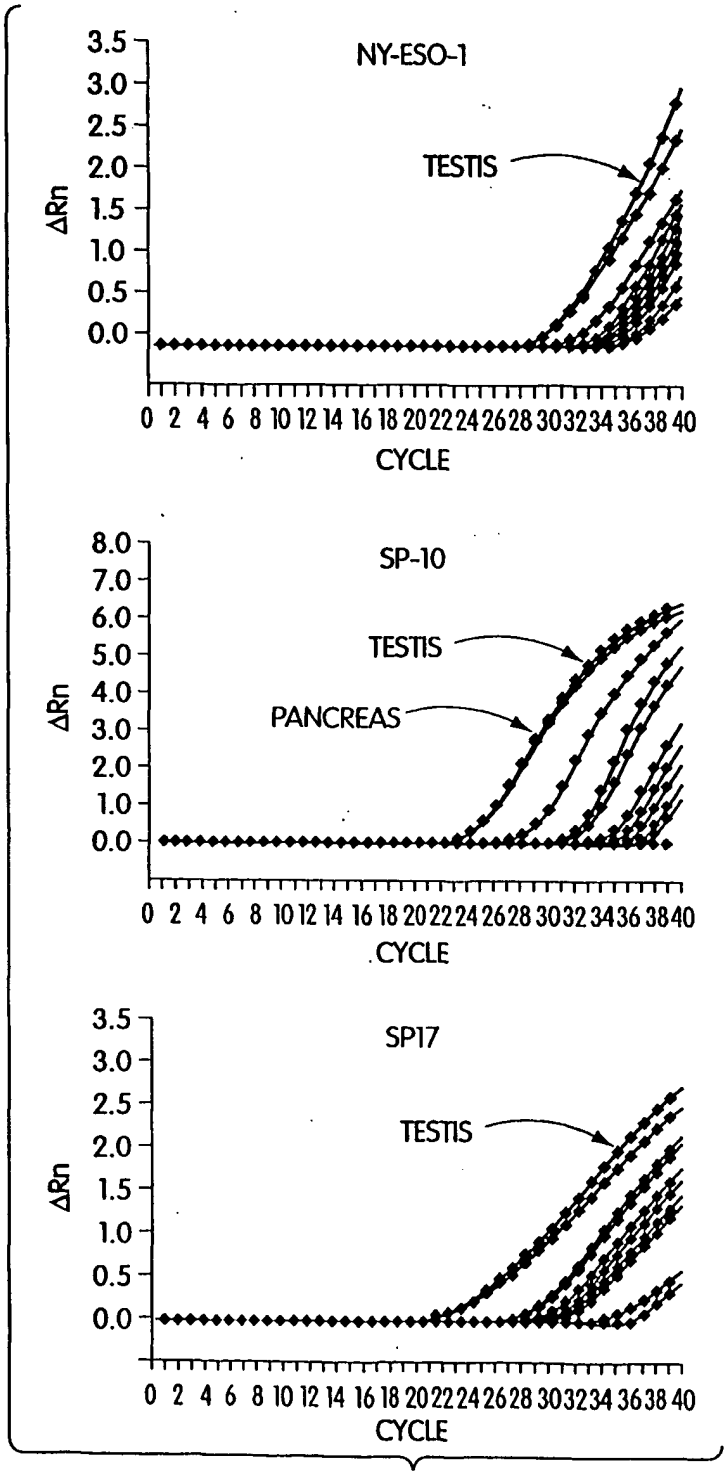

NY-ESO-1
TESTIS
ΔRn
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
CYCLE
SP-10
TESTIS
PANCREAS
ΔRn
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
CYCLE
SP17
TESTIS
ΔRn
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
CYCLE

Fig. 1A

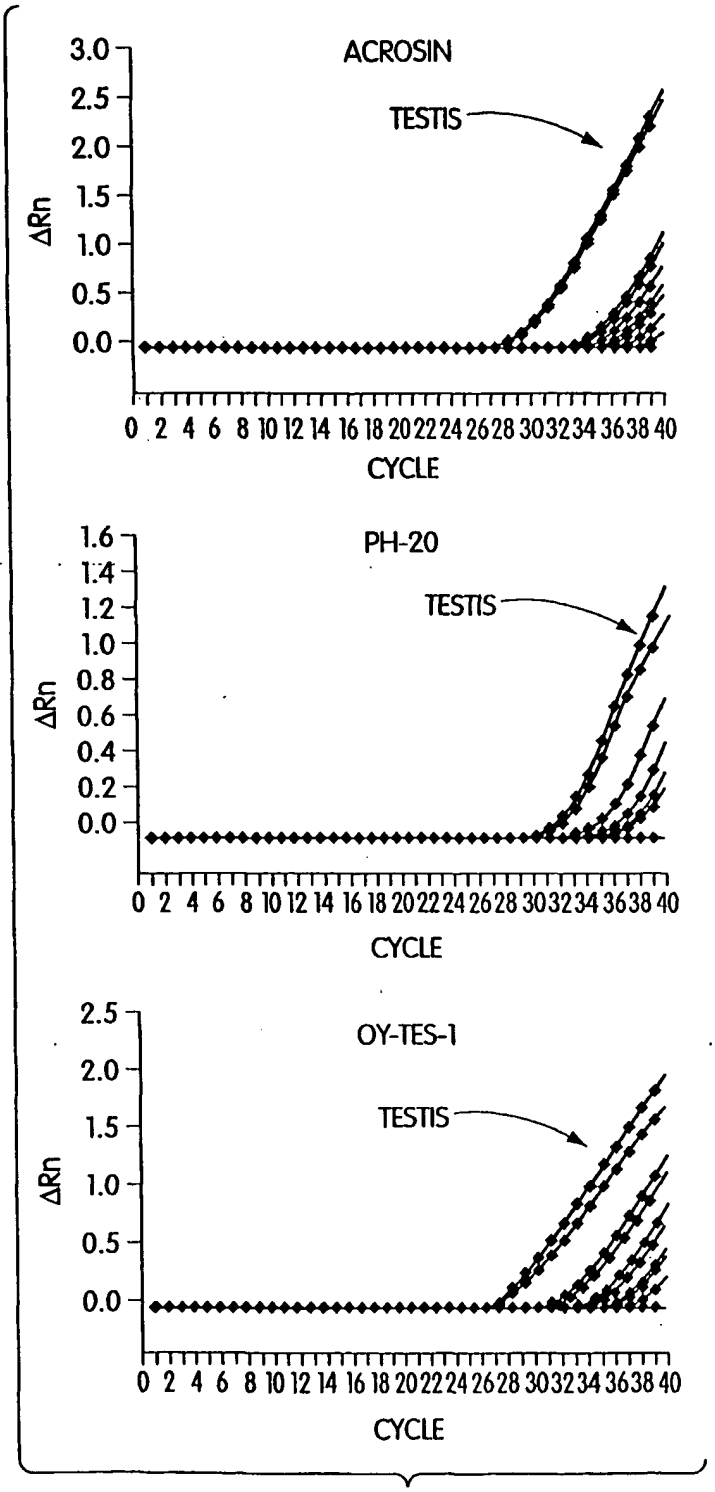



Fig. 1A
(CONTINUED)

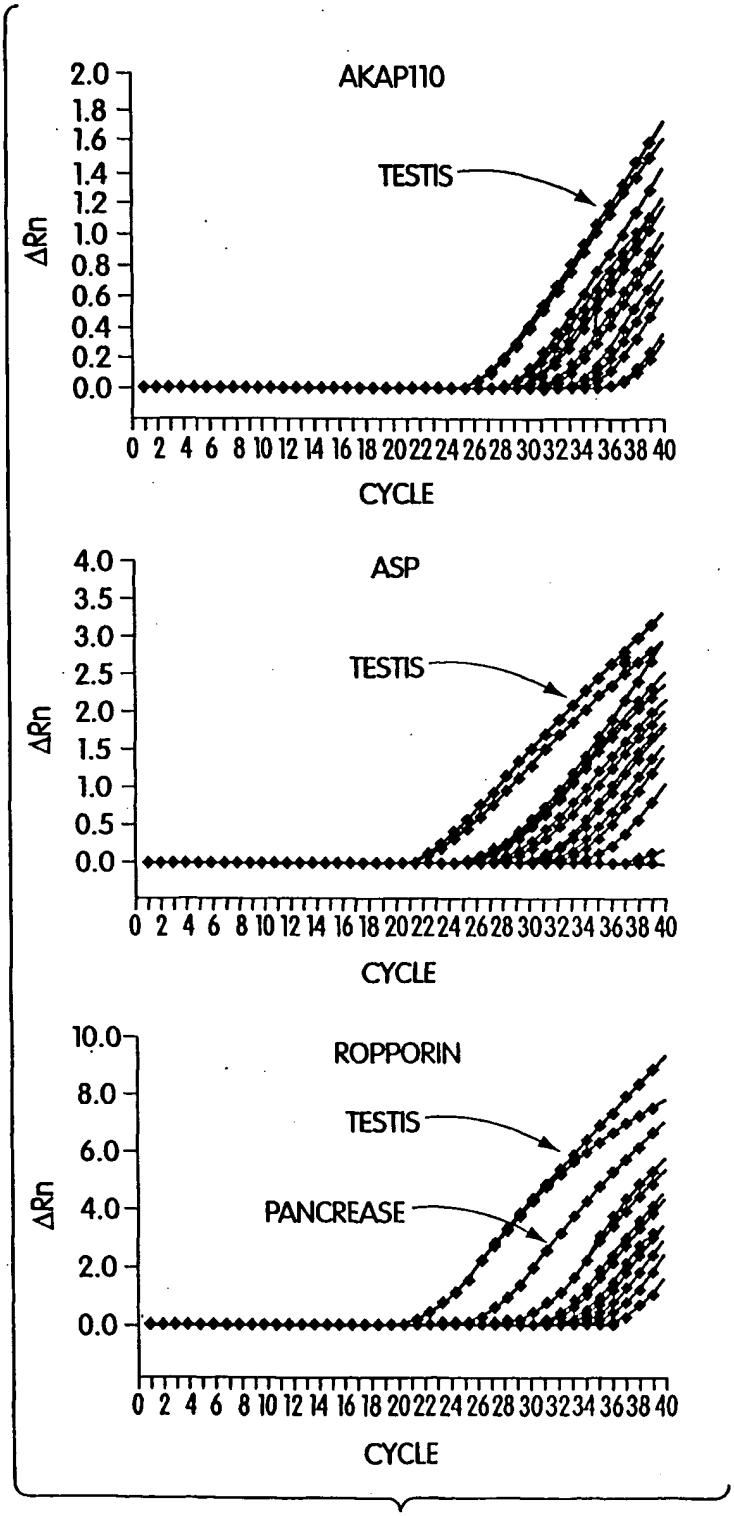



Fig. 1A
(CONTINUED)

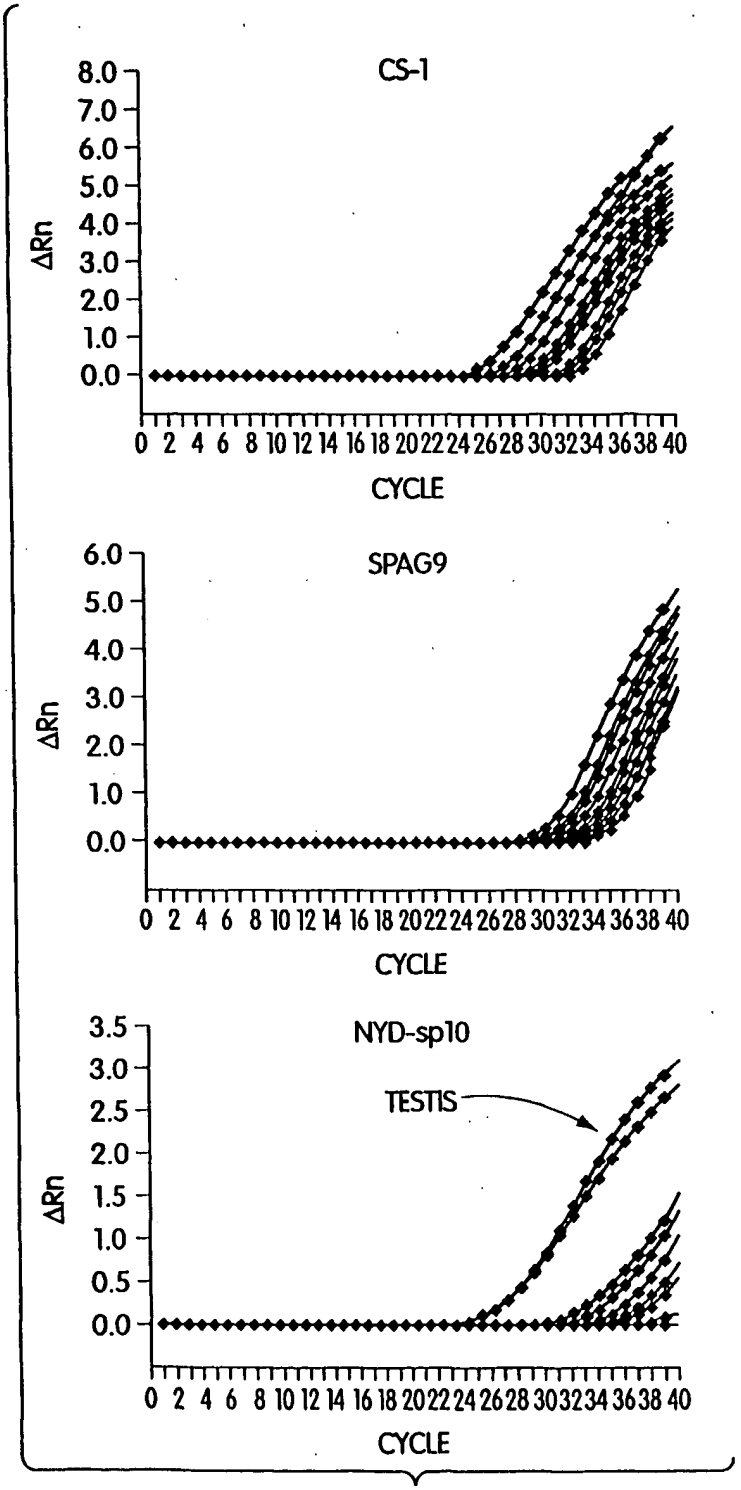



Fig. 1A
(CONTINUED)

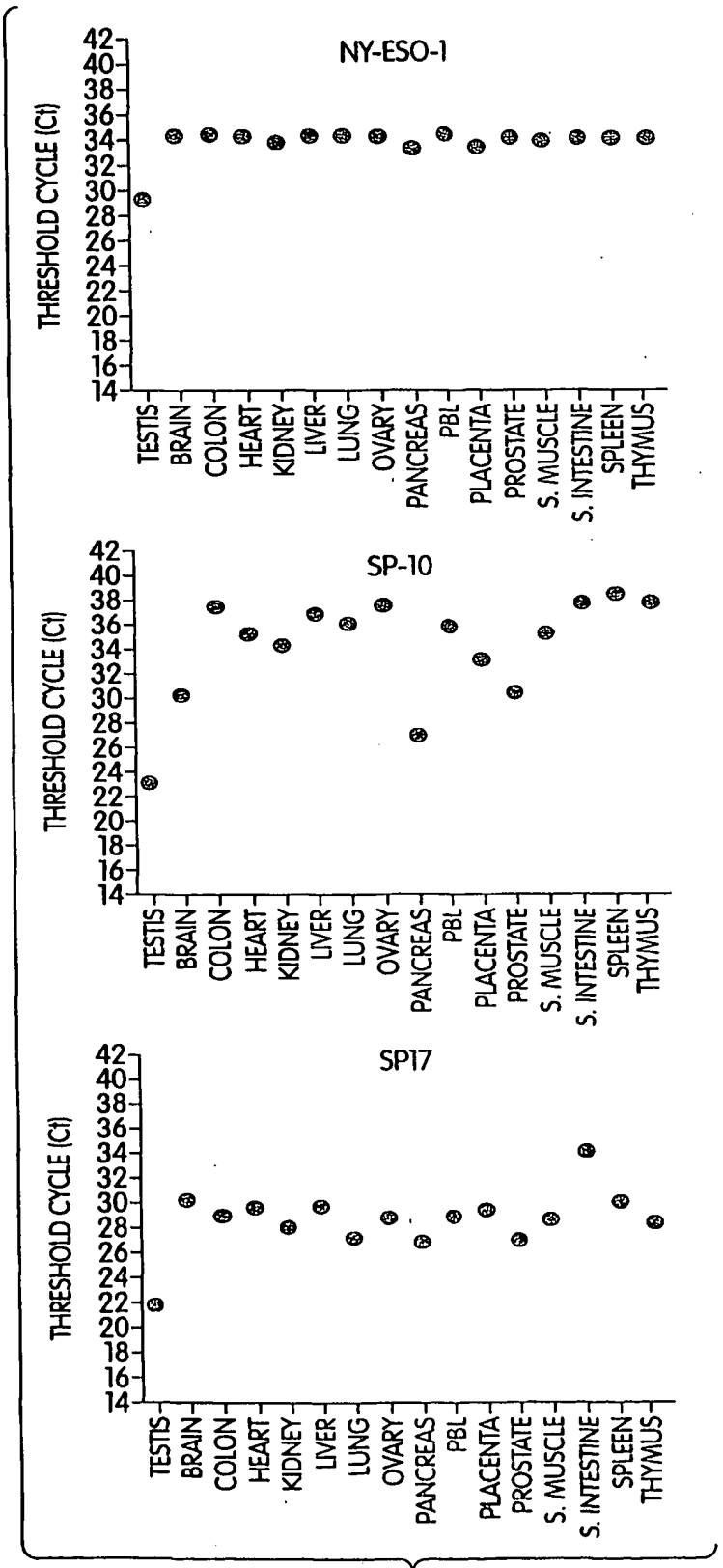

NY-ESO-1
SP-10
SP17
THRESHOLD CYCLE (Ct)
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
TESTIS
BRAIN
COLON
HEART
KIDNEY
LIVER
LUNG
OVARY
PANCREAS
PBL
PLACENTA
PROSTATE
S. MUSCLE
S. INTESTINE
SPLEEN
THYMUS

Fig. 1B

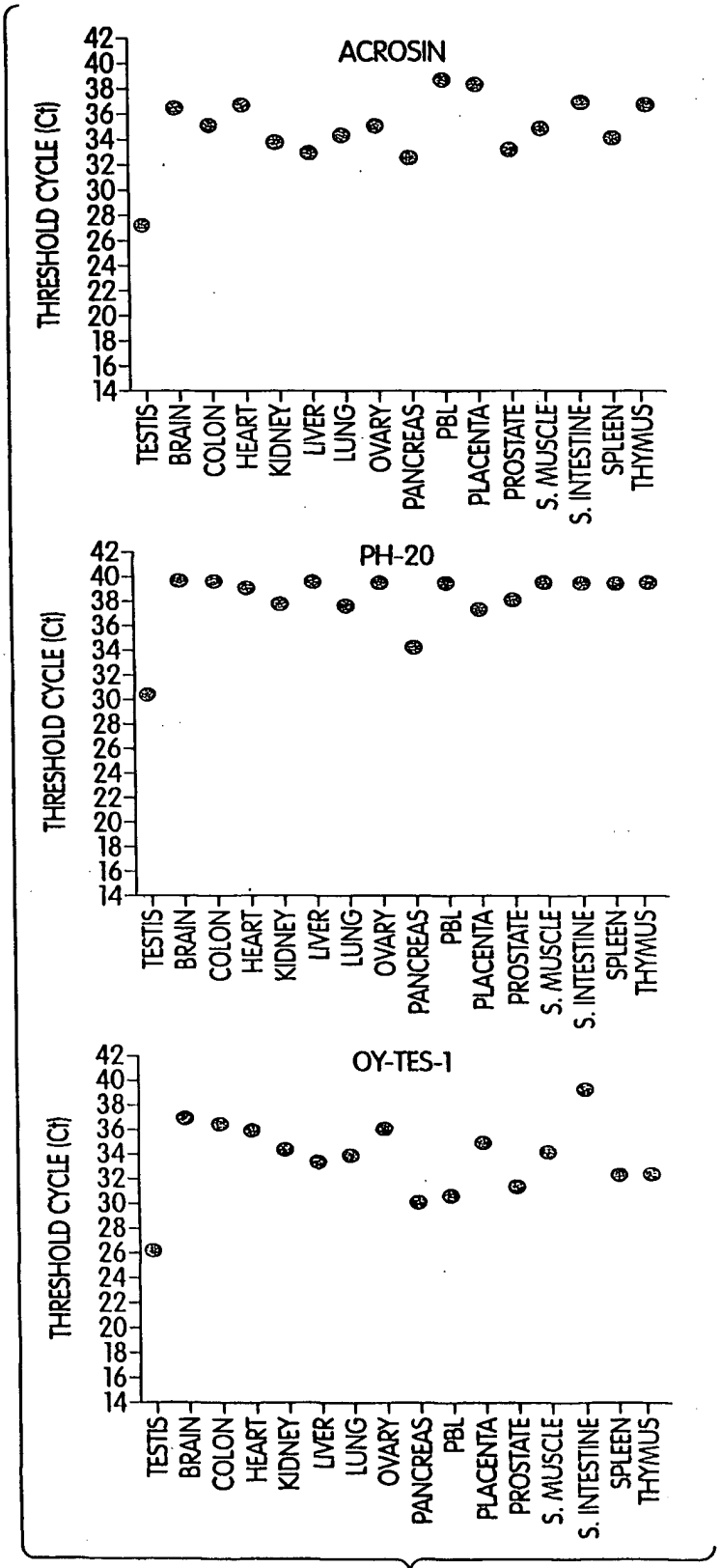



Fig. 1B
CONTINUED

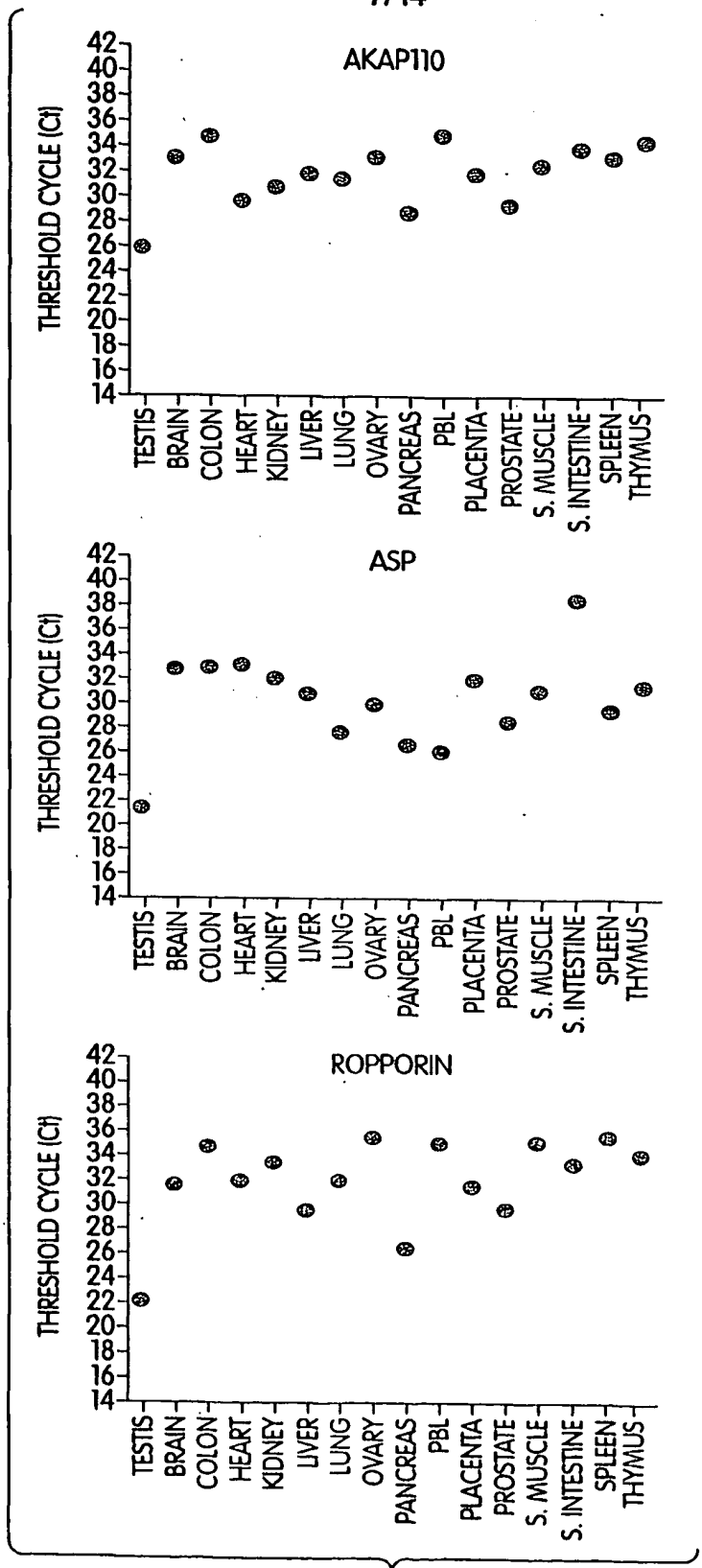



Fig. 1B
CONTINUED

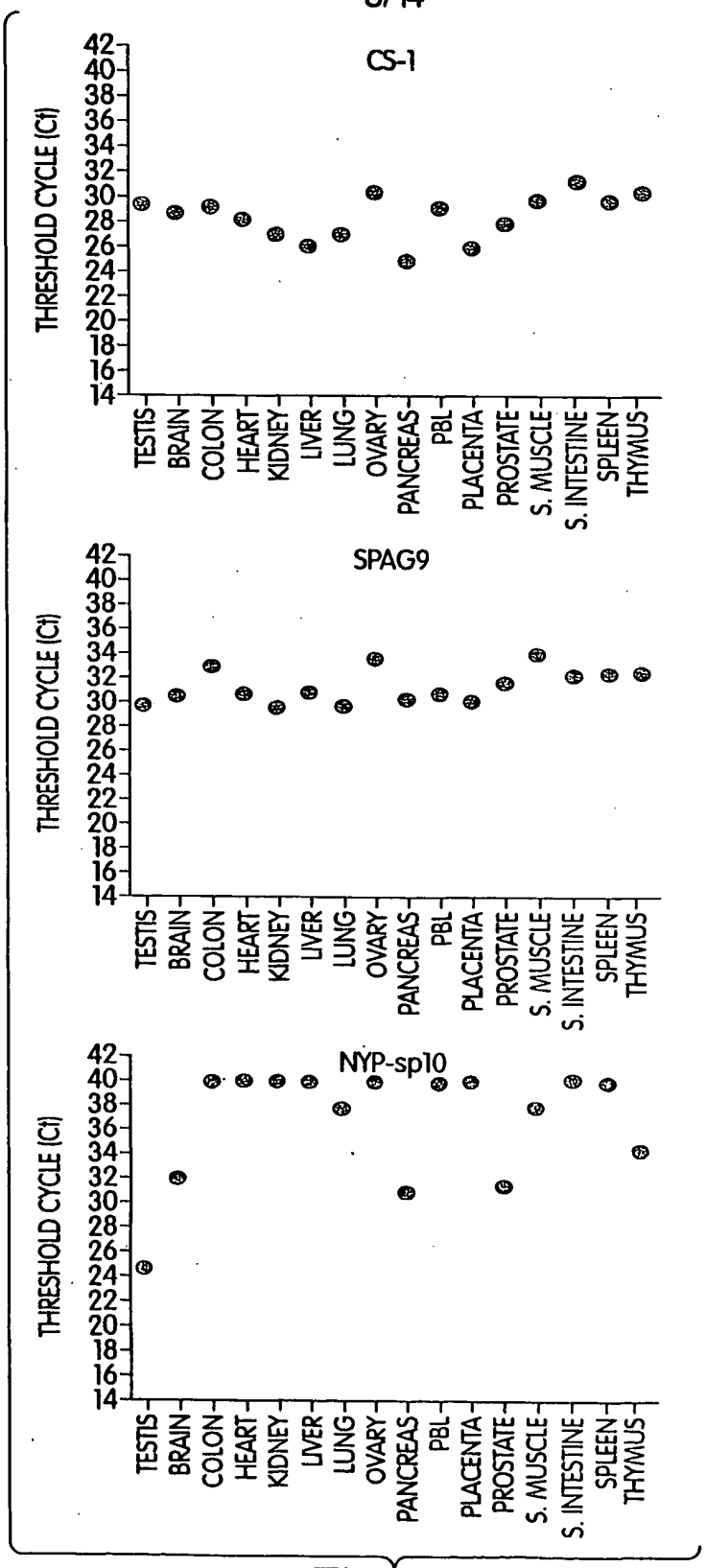



Fig. 1B
CONTINUED

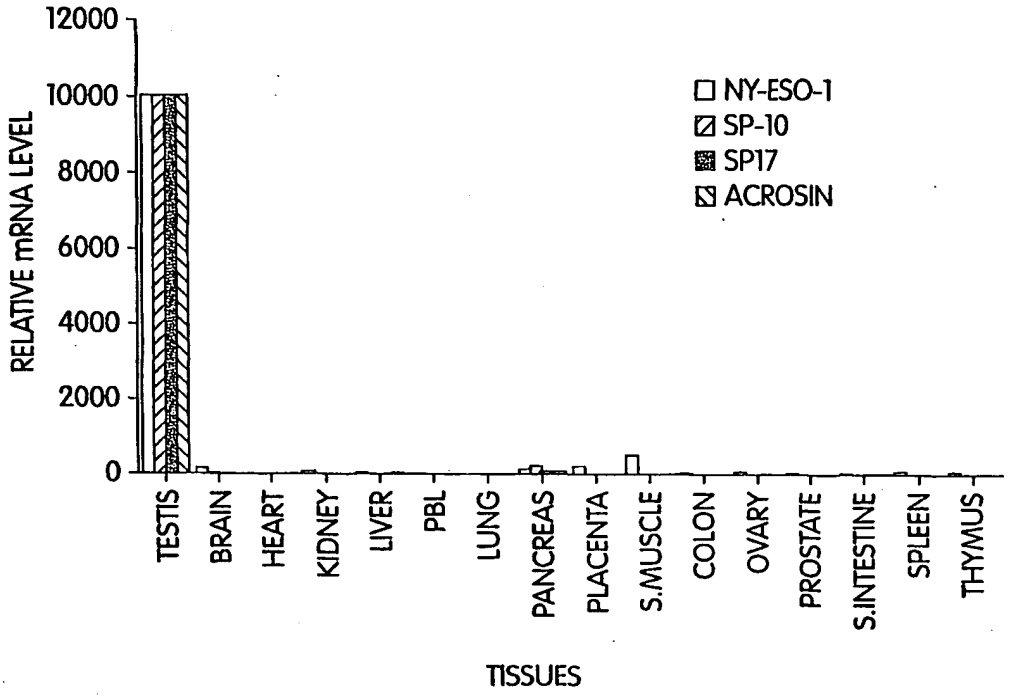

12000
10000
8000
6000
4000
2000
0
RELATIVE mRNA LEVEL
NY-ESO-1
SP-10
SP17
ACROSIN
TESTIS
BRAIN
HEART
KIDNEY
LIVER
PBL
LUNG
PANCREAS
PLACENTA
S.MUSCLE
COLON
OVARY
PROSTATE
S.INTESTINE
SPLEEN
THYMUS
TISSUES

Fig. 2A

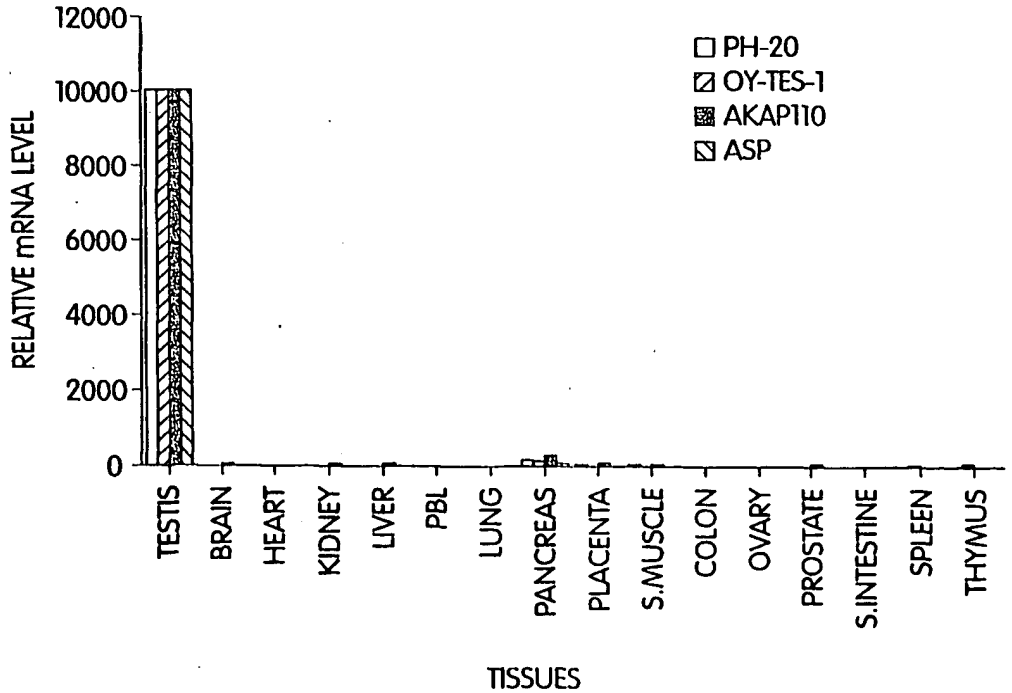

12000
10000
8000
6000
4000
2000
0
RELATIVE mRNA LEVEL
PH-20
OY-TES-1
AKAP110
ASP
TESTIS
BRAIN
HEART
KIDNEY
LIVER
PBL
LUNG
PANCREAS
PLACENTA
S.MUSCLE
COLON
OVARY
PROSTATE
S.INTESTINE
SPLEEN
THYMUS
TISSUES

Fig. 2B

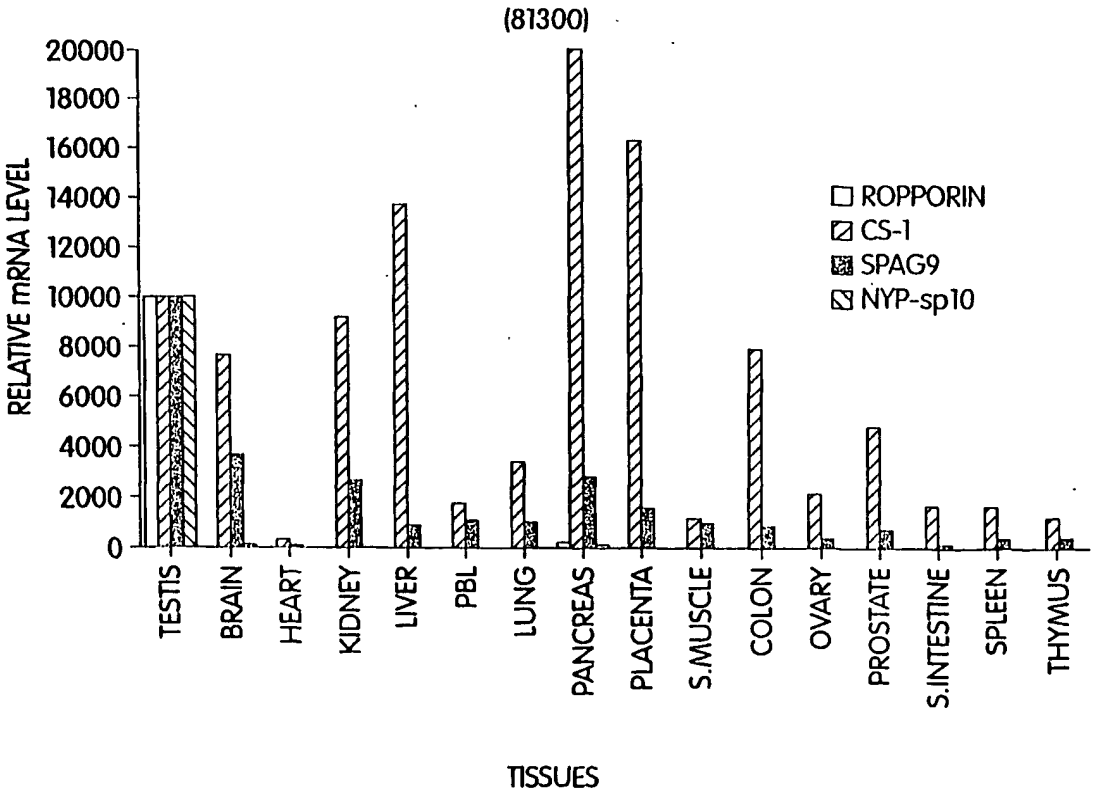

(81300)
RELATIVE mRNA LEVEL
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
ROPPORIN
CS-1
SPAG9
NYP-sp10
TESTIS
BRAIN
HEART
KIDNEY
LIVER
PBL
LUNG
PANCREAS
PLACENTA
S.MUSCLE
COLON
OVARY
PROSTATE
S.INTESTINE
SPLEEN
THYMUS
TISSUES

Fig. 2C

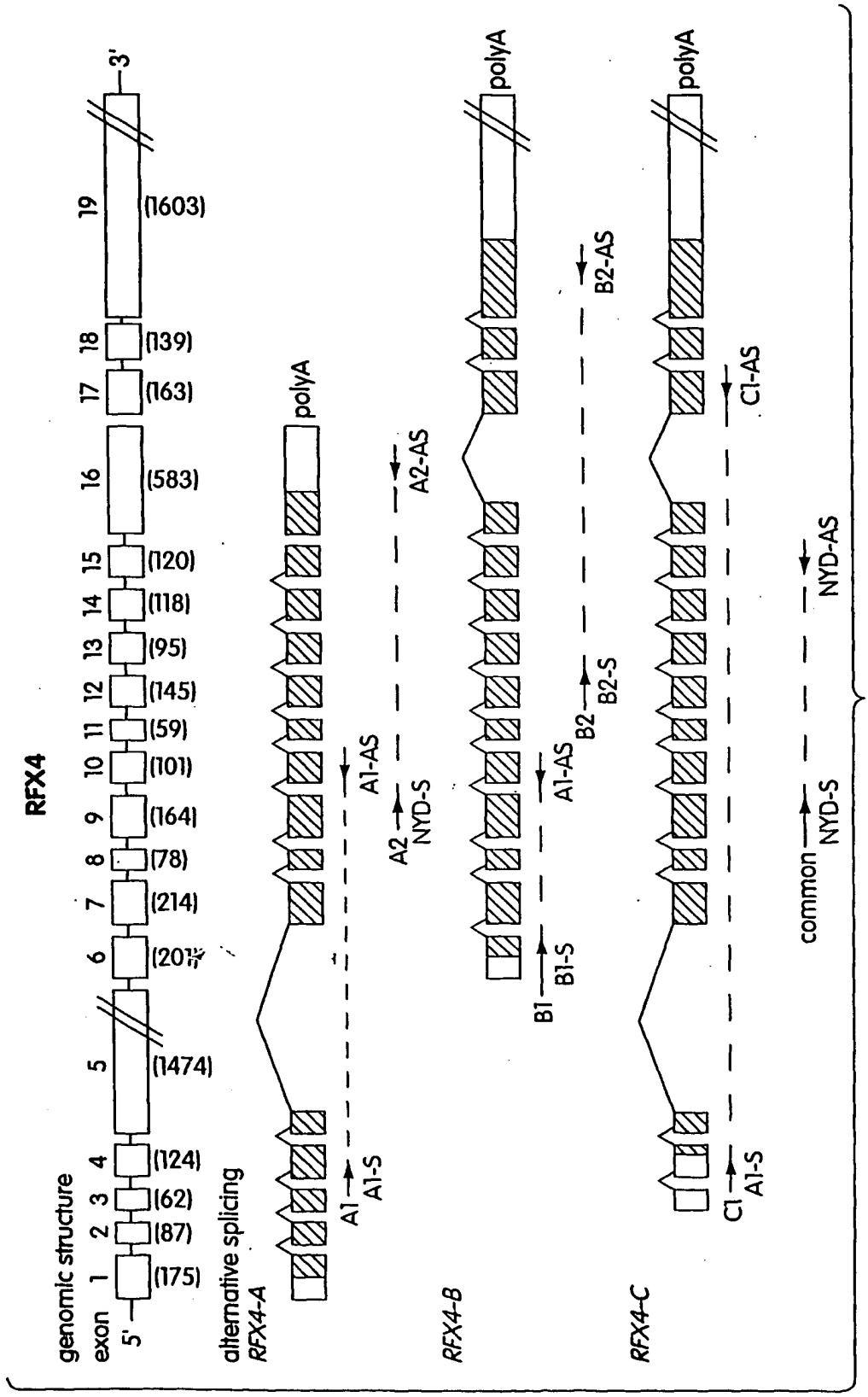

RFX4
genomic structure
exon
5'
1 (175)
2 (87)
3 (62)
4 (124)
5 (1474)
6 (201)
7 (214)
8 (78)
9 (164)
10 (101)
11 (59)
12 (145)
13 (95)
14 (118)
15 (120)
16 (583)
17 (163)
18 (139)
19 (1603)
3'
alternative splicing
RFX4-A
polyA
A1
A1-S
A1-AS
A2
NYD-S
A2-AS
RFX4-B
polyA
B1
B1-S
A1-AS
B2
B2-S
B2-AS
RFX4-C
polyA
C1
A1-S
C1-AS
common
NYD-S
NYD-AS

Fig. 3

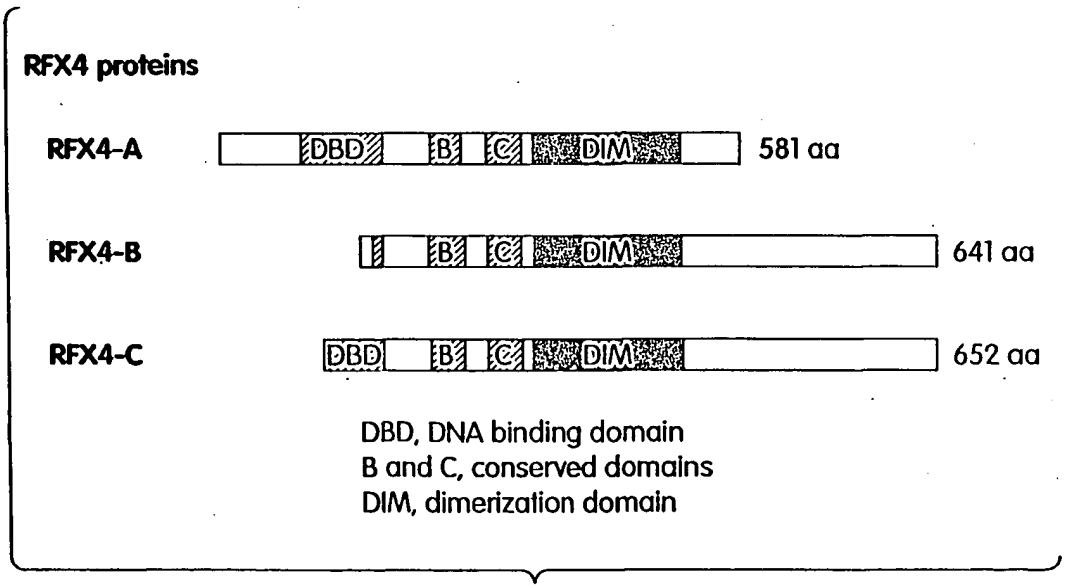

RFX4 proteins
RFX4-A
DBD
B
C
DIM
581 aa
RFX4-B
B
C
DIM
641 aa
RFX4-C
DBD
B
C
DIM
652 aa
DBD, DNA binding domain
B and C, conserved domains
DIM, dimerization domain

Fig. 4

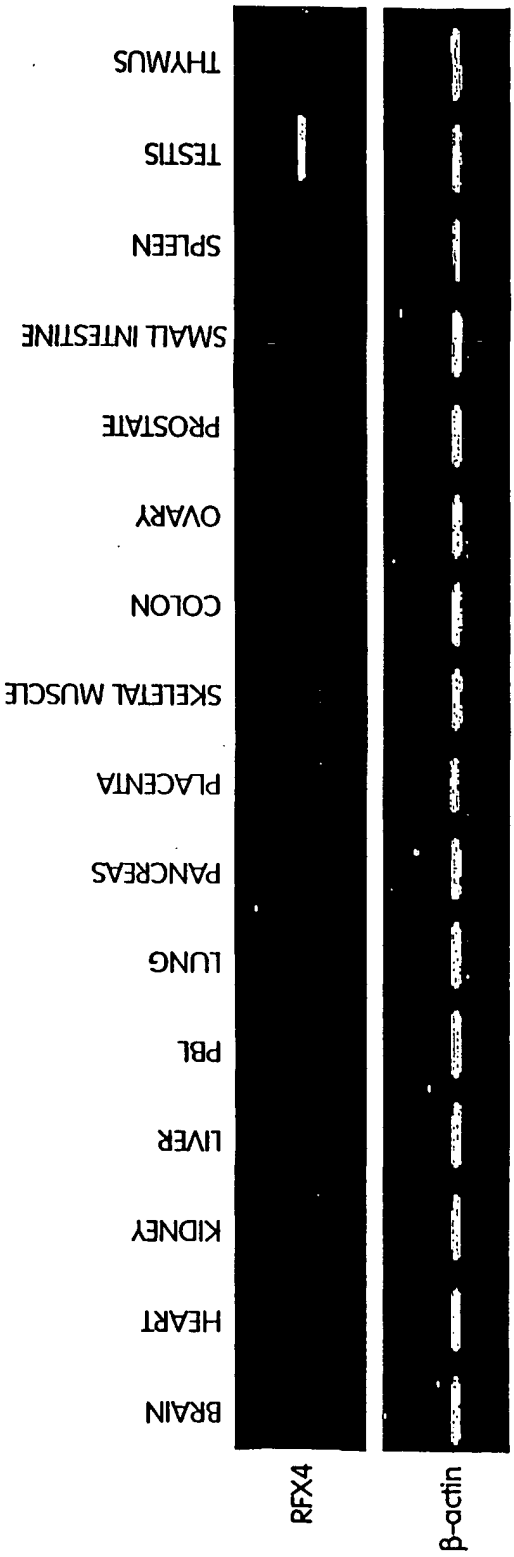

BRAIN
HEART
KIDNEY
LIVER
PBL
LUNG
PANCREAS
PLACENTA
SKELETAL MUSCLE
COLON
OVARY
PROSTATE
SMALL INTESTINE
SPLEEN
TESTIS
THYMUS
RFX4
β-actin

Fig. 5A

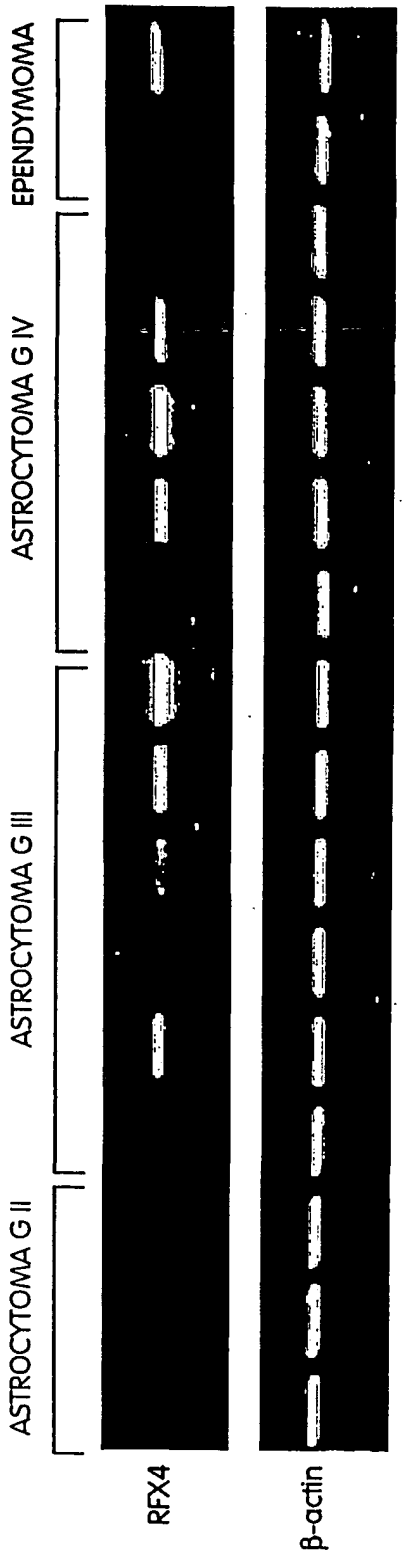

ASTROCYTOMA G II
ASTROCYTOMA G III
ASTROCYTOMA G IV
EPENDYMOMA
RFX4
β-actin

Fig. 5B

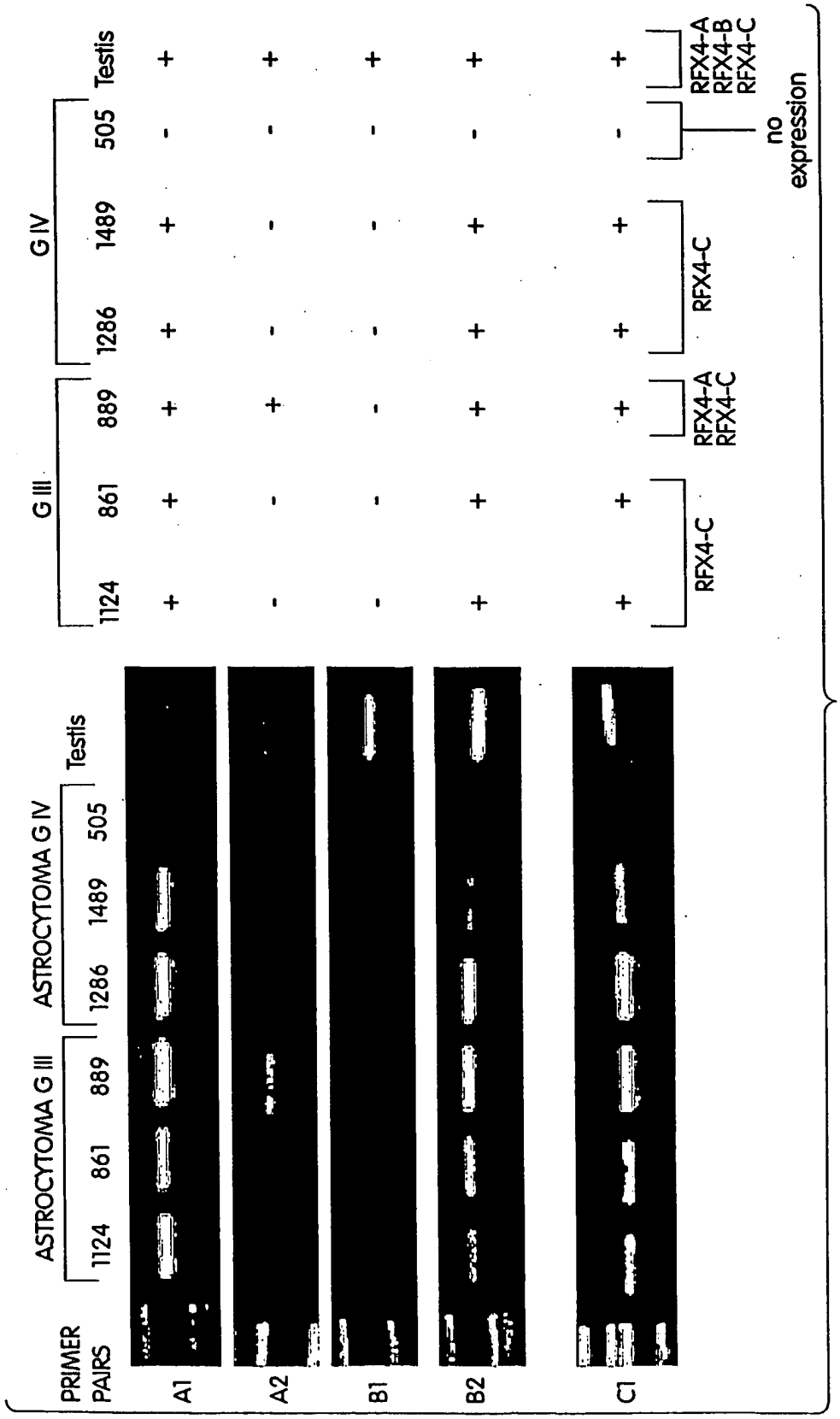

PRIMER PAIRS
ASTROCYTOMA G III
ASTROCYTOMA G IV
Testis
1124
861
889
1286
1489
505
A1
A2
B1
B2
C1
G III
G IV
RFX4-C
RFX4-A RFX4-C
RFX4-C
no expression
RFX4-A RFX4-B RFX4-C

Fig. 6

SEQUENCE LISTING

<110> Nakayama, Eiichi
Ono, Toshiro
Old, Lloyd J.

<120> CANCER-TESTIS ANTIGENS

<130> L00461/70123WO

<150> US 60/356,937
<151> 2002-02-14

<150> US 60/285,343
<151> 2001-04-20

<160> 62

<170> PatentIn version 3.1

<210> 1
<211> 1912
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (307)..(1317)
<223>

<400> 1
ctagccaatg ctctaggaag acattgagac cagccaactt cttgccttga taactactga 60

agagacattg ggtggctgga ttttgaaagc agacttctgg ttataggtga tgcaacttga 120

aaaacaatcc tgaaacatga aacaagaata ataatattta aatgtaactt aatcattata 180

cctctttatc catcaaagtg aattcattcc attccctttc atctgtgctc atactttgca 240

tcagatattg ggtaaaccaa agtgtgtagg aagaaataaa tgttttcata gtcattactc 300

tttaca atg gga gtg cta aaa ttc aag cac atc ttt ttc aga agc ttt 348
Met Gly Val Leu Lys Phe Lys His Ile Phe Phe Arg Ser Phe
1 5 10

gtt aaa tca agt gga gta tcc cag ata gtt ttc acc ttc ctt ctg att 396
Val Lys Ser Ser Gly Val Ser Gln Ile Val Phe Thr Phe Leu Leu Ile
15 20 25 30

cca tgt tgc ttg act ctg aat ttc aga gca cct cct gtt att cca aat 444
Pro Cys Cys Leu Thr Leu Asn Phe Arg Ala Pro Pro Val Ile Pro Asn
35 40 45

gtg cct ttc ctc tgg gcc tgg aat gcc cca agt gaa ttt tgt ctt gga 492
Val Pro Phe Leu Trp Ala Trp Asn Ala Pro Ser Glu Phe Cys Leu Gly
50 55 60

aaa ttt gat gag cca cta gat atg agc ctc ttc tct ttc ata gga agc 540
Lys Phe Asp Glu Pro Leu Asp Met Ser Leu Phe Ser Phe Ile Gly Ser
65 70 75

```
ccc cga ata aac gcc acc ggg caa ggt gtt aca ata ttt tat gtt gat      588
Pro Arg Ile Asn Ala Thr Gly Gln Gly Val Thr Ile Phe Tyr Val Asp
    80                  85                  90

aga ctt ggc tac tat cct tac ata gat tca atc aca gga gta act gtg      636
Arg Leu Gly Tyr Tyr Pro Tyr Ile Asp Ser Ile Thr Gly Val Thr Val
95                  100                 105                 110

aat gga gga atc ccc cag aag att tcc tta caa gac cat ctg gac aaa      684
Asn Gly Gly Ile Pro Gln Lys Ile Ser Leu Gln Asp His Leu Asp Lys
                115                 120                 125

gct aag aaa gac att aca ttt tat atg cca gta gac aat ttg gga atg      732
Ala Lys Lys Asp Ile Thr Phe Tyr Met Pro Val Asp Asn Leu Gly Met
            130                 135                 140

gct gtt att gac tgg gaa gaa tgg aga ccc act tgg gca aga aac tgg      780
Ala Val Ile Asp Trp Glu Glu Trp Arg Pro Thr Trp Ala Arg Asn Trp
        145                 150                 155

aaa cct aaa gat gtt tac aag aat agg tct att gaa ttg gtt cag caa      828
Lys Pro Lys Asp Val Tyr Lys Asn Arg Ser Ile Glu Leu Val Gln Gln
    160                 165                 170

caa aat gta caa ctt agt ctc aca gag gcc act gag aaa gca aaa caa      876
Gln Asn Val Gln Leu Ser Leu Thr Glu Ala Thr Glu Lys Ala Lys Gln
175                 180                 185                 190

gaa ttt gaa aag gca ggg aag gat ttc ctg gta gag act ata aaa ttg      924
Glu Phe Glu Lys Ala Gly Lys Asp Phe Leu Val Glu Thr Ile Lys Leu
                195                 200                 205

gga aaa tta ctt cgg cca aat cac ttg tgg ggt tat tat ctt ttt ccg      972
Gly Lys Leu Leu Arg Pro Asn His Leu Trp Gly Tyr Tyr Leu Phe Pro
            210                 215                 220

gat tgt tac aac cat cac tat aag aaa ccc ggt tac aat gga agt tgc     1020
Asp Cys Tyr Asn His His Tyr Lys Lys Pro Gly Tyr Asn Gly Ser Cys
        225                 230                 235

ttc aat gta gaa ata aaa aga aat gat gat ctc agc tgg ttg tgg aat     1068
Phe Asn Val Glu Ile Lys Arg Asn Asp Asp Leu Ser Trp Leu Trp Asn
    240                 245                 250

gaa agc act gct ctt tac cca tcc att tat ttg aac act cag cag tct     1116
Glu Ser Thr Ala Leu Tyr Pro Ser Ile Tyr Leu Asn Thr Gln Gln Ser
255                 260                 265                 270

cct gta gct gct aca ctc tat gtg cgc aat cga gtt cgg gaa gcc atc     1164
Pro Val Ala Ala Thr Leu Tyr Val Arg Asn Arg Val Arg Glu Ala Ile
                275                 280                 285

aga gtt tcc aaa ata cct gat gca aaa agt cca ctt ccg gtt ttt gca     1212
Arg Val Ser Lys Ile Pro Asp Ala Lys Ser Pro Leu Pro Val Phe Ala
            290                 295                 300

tat acc cgc ata gtt ttt act gat caa gtt ttg aaa ttc ctt tct caa     1260
Tyr Thr Arg Ile Val Phe Thr Asp Gln Val Leu Lys Phe Leu Ser Gln
        305                 310                 315
```

atg aac ttg tgt ata cat ttg gcg aaa ctg ttg ctc tgg gtg ctt ctg 1308
Met Asn Leu Cys Ile His Leu Ala Lys Leu Leu Leu Trp Val Leu Leu
320 325 330

gaa ttg taa tatggggaac cctcagtata atgcgaagta tgaaatcttg 1357
Glu Leu
335

cttgctccta gacaattaca tggagactat actgaatcct tacataatca acgtcacact 1417

agcagccaaa atgtgtagcc aagtgctttg ccaggagcaa ggagtgtgta taaggaaaaa 1477

ctggaattca agtgactatc ttcacctcaa cccagataat tttgctattc aacttgagaa 1537

aggtggaaag ttcacagtac gtggaaaacc gacacttgaa gacctggagc aatttttctga 1597

aaaattttat tgcagctgtt atagcacctt gagttgtaag gagaaagctg atgtaaaaga 1657

cactgatgct gttgatgtgt gtattgctga tggtgtctgt atagatgctt ttctaaaacc 1717

tcccatggag acagaagaac ctcaaatttt ctacaatgct tcaccctcca cactatctgc 1777

cacaatgttc attgttagta ttttgtttct tatcatttct tctgtagcga gtttgtaatt 1837

gcgcaggtta gctgaaatga acaatatgtc catcttaaag tgtgcttttt cgactaatta 1897

aatctttgaa aagaa 1912

<210> 2
<211> 336
<212> PRT
<213> Homo sapiens

<400> 2

Met Gly Val Leu Lys Phe Lys His Ile Phe Phe Arg Ser Phe Val Lys
1 5 10 15

Ser Ser Gly Val Ser Gln Ile Val Phe Thr Phe Leu Leu Ile Pro Cys
20 25 30

Cys Leu Thr Leu Asn Phe Arg Ala Pro Pro Val Ile Pro Asn Val Pro
35 40 45

Phe Leu Trp Ala Trp Asn Ala Pro Ser Glu Phe Cys Leu Gly Lys Phe
50 55 60

Asp Glu Pro Leu Asp Met Ser Leu Phe Ser Phe Ile Gly Ser Pro Arg
65 70 75 80

Ile Asn Ala Thr Gly Gln Gly Val Thr Ile Phe Tyr Val Asp Arg Leu
85 90 95

```
Gly Tyr Tyr Pro Tyr Ile Asp Ser Ile Thr Gly Val Thr Val Asn Gly
            100                 105                 110

Gly Ile Pro Gln Lys Ile Ser Leu Gln Asp His Leu Asp Lys Ala Lys
        115                 120                 125

Lys Asp Ile Thr Phe Tyr Met Pro Val Asp Asn Leu Gly Met Ala Val
    130                 135                 140

Ile Asp Trp Glu Glu Trp Arg Pro Thr Trp Ala Arg Asn Trp Lys Pro
145                 150                 155                 160

Lys Asp Val Tyr Lys Asn Arg Ser Ile Glu Leu Val Gln Gln Gln Asn
                165                 170                 175

Val Gln Leu Ser Leu Thr Glu Ala Thr Glu Lys Ala Lys Gln Glu Phe
            180                 185                 190

Glu Lys Ala Gly Lys Asp Phe Leu Val Glu Thr Ile Lys Leu Gly Lys
        195                 200                 205

Leu Leu Arg Pro Asn His Leu Trp Gly Tyr Tyr Leu Phe Pro Asp Cys
    210                 215                 220

Tyr Asn His His Tyr Lys Lys Pro Gly Tyr Asn Gly Ser Cys Phe Asn
225                 230                 235                 240

Val Glu Ile Lys Arg Asn Asp Asp Leu Ser Trp Leu Trp Asn Glu Ser
                245                 250                 255

Thr Ala Leu Tyr Pro Ser Ile Tyr Leu Asn Thr Gln Gln Ser Pro Val
            260                 265                 270

Ala Ala Thr Leu Tyr Val Arg Asn Arg Val Arg Glu Ala Ile Arg Val
        275                 280                 285

Ser Lys Ile Pro Asp Ala Lys Ser Pro Leu Pro Val Phe Ala Tyr Thr
    290                 295                 300

Arg Ile Val Phe Thr Asp Gln Val Leu Lys Phe Leu Ser Gln Met Asn
305                 310                 315                 320

Leu Cys Ile His Leu Ala Lys Leu Leu Leu Trp Val Leu Leu Glu Leu
                325                 330                 335
```

<210> 3
<211> 3014
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (230)..(2791)
<223>

<400> 3
ggtacatgga aggccacagg aagaaacaag atcttgagct gagcaagaac atcccagcat 60

cttcattgac tttaaaagta tattctggag tcttccgtgg ttcactattc cagtactaca 120

gagattcctt atattacatg gcaggagggg ggtaaactga gggatagtga agacaacaat 180

aaattaatca agagctttcc tcatatctca gaacctatcc tctgtaaga atg tca gaa 238
Met Ser Glu
1

aag gtt gac tgg tta caa agc caa aat gga gta tgc aaa gtt gat gtc 286
Lys Val Asp Trp Leu Gln Ser Gln Asn Gly Val Cys Lys Val Asp Val
5 10 15

tat tct cct gga gac aac caa gcc cag gac tgg aaa atg gac acc tcc 334
Tyr Ser Pro Gly Asp Asn Gln Ala Gln Asp Trp Lys Met Asp Thr Ser
20 25 30 35

acg gat cct gtc aga gtg ctc agc tgg ctc cgc aga gac ctg gag aag 382
Thr Asp Pro Val Arg Val Leu Ser Trp Leu Arg Arg Asp Leu Glu Lys
40 45 50

agt aca gca gag ttc caa gat gtt cgg ttc aaa ccc gga gaa tca ttt 430
Ser Thr Ala Glu Phe Gln Asp Val Arg Phe Lys Pro Gly Glu Ser Phe
55 60 65

ggt ggg gaa acg tcc aac tca gga gac cca cac aaa ggt ttc tct gta 478
Gly Gly Glu Thr Ser Asn Ser Gly Asp Pro His Lys Gly Phe Ser Val
70 75 80

gac tat tac aac acc acc acc aag ggc act cca gaa aga ttg cat ttt 526
Asp Tyr Tyr Asn Thr Thr Thr Lys Gly Thr Pro Glu Arg Leu His Phe
85 90 95

gag atg act cac aaa gag att cct tgc cag ggc ccc agg gcc caa ctt 574
Glu Met Thr His Lys Glu Ile Pro Cys Gln Gly Pro Arg Ala Gln Leu
100 105 110 115

ggc aac ggg agt tca gta gat gaa gtt tcc ttc tat gct aac cgc ctc 622
Gly Asn Gly Ser Ser Val Asp Glu Val Ser Phe Tyr Ala Asn Arg Leu
120 125 130

acg aat cta gtc ata gcc atg gcc cgc aaa gag atc aat gag aag atc 670
Thr Asn Leu Val Ile Ala Met Ala Arg Lys Glu Ile Asn Glu Lys Ile
135 140 145

gat ggc tct gaa aac aaa tgt gtc tat cag tca ttg tac atg ggg aat 718
Asp Gly Ser Glu Asn Lys Cys Val Tyr Gln Ser Leu Tyr Met Gly Asn

```
        150                 155                 160

gaa ccc aca ccc acc aaa agc ctc agt aag ata gca tca gag ctt gtg      766
Glu Pro Thr Pro Thr Lys Ser Leu Ser Lys Ile Ala Ser Glu Leu Val
    165                 170                 175

aat gag acc gtc tct gca tgt tcc agg aat gct gcc cca gac aag gct      814
Asn Glu Thr Val Ser Ala Cys Ser Arg Asn Ala Ala Pro Asp Lys Ala
180                 185                 190                 195

cct ggc tct gga gac aga gtc tca gga tca tca caa agt ccc cca aat      862
Pro Gly Ser Gly Asp Arg Val Ser Gly Ser Ser Gln Ser Pro Pro Asn
                200                 205                 210

ttg aaa tac aag tcc act ttg aag atc aag gag agc acc aaa gaa aga      910
Leu Lys Tyr Lys Ser Thr Leu Lys Ile Lys Glu Ser Thr Lys Glu Arg
            215                 220                 225

cag ggt cca gat gac aag cct cct tct aag aag tct ttc ttc tat aag      958
Gln Gly Pro Asp Asp Lys Pro Pro Ser Lys Lys Ser Phe Phe Tyr Lys
        230                 235                 240

gaa gtg ttt gaa tct cgt aac gga gat tat gcc aga gag ggt gga agg     1006
Glu Val Phe Glu Ser Arg Asn Gly Asp Tyr Ala Arg Glu Gly Gly Arg
    245                 250                 255

ttc ttt cct cgg gag aga aag agg ttt cga ggg cag gaa agg cct gat     1054
Phe Phe Pro Arg Glu Arg Lys Arg Phe Arg Gly Gln Glu Arg Pro Asp
260                 265                 270                 275

gac ttt acg gct tct gtt agt gaa ggg atc atg acc tat gct aac agt     1102
Asp Phe Thr Ala Ser Val Ser Glu Gly Ile Met Thr Tyr Ala Asn Ser
                280                 285                 290

gtg gta tct gat atg atg gtc tcc atc atg aag aca ctg aag atc caa     1150
Val Val Ser Asp Met Met Val Ser Ile Met Lys Thr Leu Lys Ile Gln
            295                 300                 305

gtg aaa gac aca acc att gcc acc atc cta ctg aag aag gtt ctg ctc     1198
Val Lys Asp Thr Thr Ile Ala Thr Ile Leu Leu Lys Lys Val Leu Leu
        310                 315                 320

aag cat gca aaa gag gtg gtc tcg gat ctc atc gac tcc ttc ttg agg     1246
Lys His Ala Lys Glu Val Val Ser Asp Leu Ile Asp Ser Phe Leu Arg
    325                 330                 335

aat ctc cac agc gtc aca ggg acc ctc atg act gac aca cag ttt gtc     1294
Asn Leu His Ser Val Thr Gly Thr Leu Met Thr Asp Thr Gln Phe Val
340                 345                 350                 355

tcg gct gtg aaa aga act gtc ttc tct cat gga agc caa aag gcc aca     1342
Ser Ala Val Lys Arg Thr Val Phe Ser His Gly Ser Gln Lys Ala Thr
                360                 365                 370

gat atc atg gat gcc atg cta agg aag ctg tac aat gta atg ttt gcc     1390
Asp Ile Met Asp Ala Met Leu Arg Lys Leu Tyr Asn Val Met Phe Ala
            375                 380                 385

aag aaa gtc cct gag cat gtc agg aaa gcc caa gac aag gct gag agt     1438
Lys Lys Val Pro Glu His Val Arg Lys Ala Gln Asp Lys Ala Glu Ser
```

```
        390                 395                 400

tat tcc ctc atc tcc atg aaa gga atg ggt gat cct aaa aac cga aat     1486
Tyr Ser Leu Ile Ser Met Lys Gly Met Gly Asp Pro Lys Asn Arg Asn
    405                 410                 415

gtg aac ttt gcc atg aaa tct gaa act aaa ttg aga gaa aaa atg tat     1534
Val Asn Phe Ala Met Lys Ser Glu Thr Lys Leu Arg Glu Lys Met Tyr
420                 425                 430                 435

tct gaa ccc aaa tca gag gag gag act tgt gcg aaa act ctg ggt gag     1582
Ser Glu Pro Lys Ser Glu Glu Glu Thr Cys Ala Lys Thr Leu Gly Glu
                440                 445                 450

cac att atc aaa gag ggg ctt acc ctg tgg cat aaa agt cag cag aaa     1630
His Ile Ile Lys Glu Gly Leu Thr Leu Trp His Lys Ser Gln Gln Lys
            455                 460                 465

gaa tgt aaa tct cta ggt ttc cag cat gca gca ttc gaa gct ccc aac     1678
Glu Cys Lys Ser Leu Gly Phe Gln His Ala Ala Phe Glu Ala Pro Asn
        470                 475                 480

aca cag cgt aag cct gca tca gac att tcc ttt gag tac cct gaa gat     1726
Thr Gln Arg Lys Pro Ala Ser Asp Ile Ser Phe Glu Tyr Pro Glu Asp
    485                 490                 495

att ggc aac ctc agc ctt cct cca tat cct cca gag aaa cct gag aat     1774
Ile Gly Asn Leu Ser Leu Pro Pro Tyr Pro Pro Glu Lys Pro Glu Asn
500                 505                 510                 515

ttt atg tat gat tca gac tcc tgg gcc aag gac ctg atc gtg tct gcc     1822
Phe Met Tyr Asp Ser Asp Ser Trp Ala Lys Asp Leu Ile Val Ser Ala
                520                 525                 530

ctg ctt ctg att caa tat cac ctg gcc cag gga gga aga agg gat gca     1870
Leu Leu Leu Ile Gln Tyr His Leu Ala Gln Gly Gly Arg Arg Asp Ala
            535                 540                 545

cgg agc ttc gtt gaa gct gct ggc acc acc aac ttt cct gcc aat gaa     1918
Arg Ser Phe Val Glu Ala Ala Gly Thr Thr Asn Phe Pro Ala Asn Glu
        550                 555                 560

cct cct gta gct ccc gat gaa tct tgc ctt aag tct gct ccc att gta     1966
Pro Pro Val Ala Pro Asp Glu Ser Cys Leu Lys Ser Ala Pro Ile Val
    565                 570                 575

ggt gac caa gaa caa gca gaa aag aag gac cta agg agt gtt ttc ttt     2014
Gly Asp Gln Glu Gln Ala Glu Lys Lys Asp Leu Arg Ser Val Phe Phe
580                 585                 590                 595

aat ttc atc cgg aac tta ctt agt gag acc att ttc aag cgt gac cag     2062
Asn Phe Ile Arg Asn Leu Leu Ser Glu Thr Ile Phe Lys Arg Asp Gln
                600                 605                 610

agc cct gaa ccc aag gtg ccg gaa cag cca gtt aag gaa gat agg aag     2110
Ser Pro Glu Pro Lys Val Pro Glu Gln Pro Val Lys Glu Asp Arg Lys
            615                 620                 625

ttg tgt gaa aga ccg ttg gcg tct tct ccc ccc agg cta tat gag gat     2158
Leu Cys Glu Arg Pro Leu Ala Ser Ser Pro Pro Arg Leu Tyr Glu Asp
```

```
        630                 635                 640

gat gag acc cct ggt gcc ctt tct ggg ctg acc aag atg gct gtc agc     2206
Asp Glu Thr Pro Gly Ala Leu Ser Gly Leu Thr Lys Met Ala Val Ser
    645                 650                 655

cag ata gat ggc cac atg agt ggg cag atg gta gaa cat ctg atg aac     2254
Gln Ile Asp Gly His Met Ser Gly Gln Met Val Glu His Leu Met Asn
660                 665                 670                 675

tca gtg atg aag ctg tgt gtc atc att gct aag tcc tgt gat gct tcg     2302
Ser Val Met Lys Leu Cys Val Ile Ile Ala Lys Ser Cys Asp Ala Ser
                680                 685                 690

ttg gca gag ctg gga gat gac aag tct gga gat gcc agt agg cta act     2350
Leu Ala Glu Leu Gly Asp Asp Lys Ser Gly Asp Ala Ser Arg Leu Thr
            695                 700                 705

tcg gcc ttc cca gat agt tta tat gag tgc tta cca gcc aag ggc aca     2398
Ser Ala Phe Pro Asp Ser Leu Tyr Glu Cys Leu Pro Ala Lys Gly Thr
        710                 715                 720

ggg tca gca gaa gct gtc ctg cag aat gcc tat caa gct atc cat aat     2446
Gly Ser Ala Glu Ala Val Leu Gln Asn Ala Tyr Gln Ala Ile His Asn
    725                 730                 735

gaa atg aga ggc aca tca gga cag ccc cct gaa ggg tgt gca gca ccc     2494
Glu Met Arg Gly Thr Ser Gly Gln Pro Pro Glu Gly Cys Ala Ala Pro
740                 745                 750                 755

acg gtg att gtc agc aat cac aac cta acg gac aca gtt cag aac aag     2542
Thr Val Ile Val Ser Asn His Asn Leu Thr Asp Thr Val Gln Asn Lys
                760                 765                 770

caa ctc caa gcc gtc ctt caa tgg gta gct gcc tct gag ctc aat gtc     2590
Gln Leu Gln Ala Val Leu Gln Trp Val Ala Ala Ser Glu Leu Asn Val
            775                 780                 785

cct att ttg tat ttt gct ggt gat gat gaa ggg atc cag gag aag cta     2638
Pro Ile Leu Tyr Phe Ala Gly Asp Asp Glu Gly Ile Gln Glu Lys Leu
        790                 795                 800

ctt cag ctc tca gct gct gct gtg gac aaa gga tgc agt gtg ggc gag     2686
Leu Gln Leu Ser Ala Ala Ala Val Asp Lys Gly Cys Ser Val Gly Glu
    805                 810                 815

gtt ctg cag tcg gtg ctg cgc tat gag aag gag cgc cag ctg aat gag     2734
Val Leu Gln Ser Val Leu Arg Tyr Glu Lys Glu Arg Gln Leu Asn Glu
820                 825                 830                 835

gcg gtg ggg aat gtc aca ccg ctg cag ctg ctg gac tgg ctg atg gtg     2782
Ala Val Gly Asn Val Thr Pro Leu Gln Leu Leu Asp Trp Leu Met Val
                840                 845                 850

aac ctg taa tcggcaaccc cactgctttc ccctcttctg gcagtggggc             2831
Asn Leu

cggcccttat ccccgccctt ctttctcact tccacatctc cccctctata tcctcacaga   2891
```

gccctaacat tatcttcaca ccactctcat caaagacatg tcatcttgtg ctagccactg 2951

gattttgcag attttcctgt ccatgcaagc aaggacgtaa aattaaaaaa ttacaattaa 3011

aaa 3014

<210> 4
<211> 853
<212> PRT
<213> Homo sapiens

<400> 4

Met Ser Glu Lys Val Asp Trp Leu Gln Ser Gln Asn Gly Val Cys Lys
1 5 10 15

Val Asp Val Tyr Ser Pro Gly Asp Asn Gln Ala Gln Asp Trp Lys Met
20 25 30

Asp Thr Ser Thr Asp Pro Val Arg Val Leu Ser Trp Leu Arg Arg Asp
35 40 45

Leu Glu Lys Ser Thr Ala Glu Phe Gln Asp Val Arg Phe Lys Pro Gly
50 55 60

Glu Ser Phe Gly Gly Glu Thr Ser Asn Ser Gly Asp Pro His Lys Gly
65 70 75 80

Phe Ser Val Asp Tyr Tyr Asn Thr Thr Thr Lys Gly Thr Pro Glu Arg
85 90 95

Leu His Phe Glu Met Thr His Lys Glu Ile Pro Cys Gln Gly Pro Arg
100 105 110

Ala Gln Leu Gly Asn Gly Ser Ser Val Asp Glu Val Ser Phe Tyr Ala
115 120 125

Asn Arg Leu Thr Asn Leu Val Ile Ala Met Ala Arg Lys Glu Ile Asn
130 135 140

Glu Lys Ile Asp Gly Ser Glu Asn Lys Cys Val Tyr Gln Ser Leu Tyr
145 150 155 160

Met Gly Asn Glu Pro Thr Pro Thr Lys Ser Leu Ser Lys Ile Ala Ser
165 170 175

Glu Leu Val Asn Glu Thr Val Ser Ala Cys Ser Arg Asn Ala Ala Pro
180 185 190

Asp Lys Ala Pro Gly Ser Gly Asp Arg Val Ser Gly Ser Ser Gln Ser
195 200 205

Pro Pro Asn Leu Lys Tyr Lys Ser Thr Leu Lys Ile Lys Glu Ser Thr
210 215 220

Lys Glu Arg Gln Gly Pro Asp Asp Lys Pro Pro Ser Lys Lys Ser Phe
225 230 235 240

Phe Tyr Lys Glu Val Phe Glu Ser Arg Asn Gly Asp Tyr Ala Arg Glu
245 250 255

Gly Gly Arg Phe Phe Pro Arg Glu Arg Lys Arg Phe Arg Gly Gln Glu
260 265 270

Arg Pro Asp Asp Phe Thr Ala Ser Val Ser Glu Gly Ile Met Thr Tyr
275 280 285

Ala Asn Ser Val Val Ser Asp Met Met Val Ser Ile Met Lys Thr Leu
290 295 300

Lys Ile Gln Val Lys Asp Thr Thr Ile Ala Thr Ile Leu Leu Lys Lys
305 310 315 320

Val Leu Leu Lys His Ala Lys Glu Val Val Ser Asp Leu Ile Asp Ser
325 330 335

Phe Leu Arg Asn Leu His Ser Val Thr Gly Thr Leu Met Thr Asp Thr
340 345 350

Gln Phe Val Ser Ala Val Lys Arg Thr Val Phe Ser His Gly Ser Gln
355 360 365

Lys Ala Thr Asp Ile Met Asp Ala Met Leu Arg Lys Leu Tyr Asn Val
370 375 380

Met Phe Ala Lys Lys Val Pro Glu His Val Arg Lys Ala Gln Asp Lys
385 390 395 400

Ala Glu Ser Tyr Ser Leu Ile Ser Met Lys Gly Met Gly Asp Pro Lys
405 410 415

Asn Arg Asn Val Asn Phe Ala Met Lys Ser Glu Thr Lys Leu Arg Glu
420 425 430

Lys Met Tyr Ser Glu Pro Lys Ser Glu Glu Glu Thr Cys Ala Lys Thr
435 440 445

Leu Gly Glu His Ile Ile Lys Glu Gly Leu Thr Leu Trp His Lys Ser
450 455 460

Gln Gln Lys Glu Cys Lys Ser Leu Gly Phe Gln His Ala Ala Phe Glu
465 470 475 480

Ala Pro Asn Thr Gln Arg Lys Pro Ala Ser Asp Ile Ser Phe Glu Tyr
485 490 495

Pro Glu Asp Ile Gly Asn Leu Ser Leu Pro Pro Tyr Pro Pro Glu Lys
500 505 510

Pro Glu Asn Phe Met Tyr Asp Ser Asp Ser Trp Ala Lys Asp Leu Ile
515 520 525

Val Ser Ala Leu Leu Leu Ile Gln Tyr His Leu Ala Gln Gly Gly Arg
530 535 540

Arg Asp Ala Arg Ser Phe Val Glu Ala Ala Gly Thr Thr Asn Phe Pro
545 550 555 560

Ala Asn Glu Pro Pro Val Ala Pro Asp Glu Ser Cys Leu Lys Ser Ala
565 570 575

Pro Ile Val Gly Asp Gln Glu Gln Ala Glu Lys Lys Asp Leu Arg Ser
580 585 590

Val Phe Phe Asn Phe Ile Arg Asn Leu Leu Ser Glu Thr Ile Phe Lys
595 600 605

Arg Asp Gln Ser Pro Glu Pro Lys Val Pro Glu Gln Pro Val Lys Glu
610 615 620

Asp Arg Lys Leu Cys Glu Arg Pro Leu Ala Ser Ser Pro Pro Arg Leu
625 630 635 640

Tyr Glu Asp Asp Glu Thr Pro Gly Ala Leu Ser Gly Leu Thr Lys Met
645 650 655

Ala Val Ser Gln Ile Asp Gly His Met Ser Gly Gln Met Val Glu His
660 665 670

Leu Met Asn Ser Val Met Lys Leu Cys Val Ile Ile Ala Lys Ser Cys
675 680 685

Asp Ala Ser Leu Ala Glu Leu Gly Asp Asp Lys Ser Gly Asp Ala Ser
690 695 700

Arg Leu Thr Ser Ala Phe Pro Asp Ser Leu Tyr Glu Cys Leu Pro Ala
705 710 715 720

Lys Gly Thr Gly Ser Ala Glu Ala Val Leu Gln Asn Ala Tyr Gln Ala
725 730 735

Ile His Asn Glu Met Arg Gly Thr Ser Gly Gln Pro Pro Glu Gly Cys
740 745 750

Ala Ala Pro Thr Val Ile Val Ser Asn His Asn Leu Thr Asp Thr Val
755 760 765

Gln Asn Lys Gln Leu Gln Ala Val Leu Gln Trp Val Ala Ala Ser Glu
770 775 780

Leu Asn Val Pro Ile Leu Tyr Phe Ala Gly Asp Asp Glu Gly Ile Gln
785 790 795 800

Glu Lys Leu Leu Gln Leu Ser Ala Ala Ala Val Asp Lys Gly Cys Ser
805 810 815

Val Gly Glu Val Leu Gln Ser Val Leu Arg Tyr Glu Lys Glu Arg Gln
820 825 830

Leu Asn Glu Ala Val Gly Asn Val Thr Pro Leu Gln Leu Leu Asp Trp
835 840 845

Leu Met Val Asn Leu
850

<210> 5
<211> 1375
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (18)..(1283)
<223>

<400> 5

```
caggcagtgc caggagt atg gtt gag atg cta cca act gcc att ctg ctg          50
                  Met Val Glu Met Leu Pro Thr Ala Ile Leu Leu
                  1               5                   10

gtc ttg gca gtg tcc gtg gtt gct aaa gat aac gcc acg tgt gat ggc          98
Val Leu Ala Val Ser Val Val Ala Lys Asp Asn Ala Thr Cys Asp Gly
            15                  20                  25

ccc tgt ggg tta cgg ttc agg caa aac cca cag ggt ggt gtc cgc atc         146
Pro Cys Gly Leu Arg Phe Arg Gln Asn Pro Gln Gly Gly Val Arg Ile
        30                  35                  40

gtc ggc ggg aag gct gca cag cat ggg gcc tgg ccc tgg atg gtc agc         194
Val Gly Gly Lys Ala Ala Gln His Gly Ala Trp Pro Trp Met Val Ser
    45                  50                  55

ctc cag atc ttc acg tac aac agc cac agg tac cac aca tgt gga ggc         242
Leu Gln Ile Phe Thr Tyr Asn Ser His Arg Tyr His Thr Cys Gly Gly
60                  65                  70                  75

agc ttg ctg aat tca cga tgg gtg ctc act gct gct cac tgc ttc gtc         290
Ser Leu Leu Asn Ser Arg Trp Val Leu Thr Ala Ala His Cys Phe Val
                80                  85                  90

ggc aaa aat aat gtg cat gac tgg aga ctg gtt ttc gga gca aag gaa         338
Gly Lys Asn Asn Val His Asp Trp Arg Leu Val Phe Gly Ala Lys Glu
            95                  100                 105

att aca tat ggg aac aat aaa cca gta aag gcg cct ctg caa gag aga         386
Ile Thr Tyr Gly Asn Asn Lys Pro Val Lys Ala Pro Leu Gln Glu Arg
        110                 115                 120

tat gtg gag aaa atc atc att cat gaa aaa tac aac tct gcg aca gag         434
Tyr Val Glu Lys Ile Ile Ile His Glu Lys Tyr Asn Ser Ala Thr Glu
    125                 130                 135

gga aat gac att gcc ctc gtg gag atc acc cct ccc att tcg tgt ggg         482
Gly Asn Asp Ile Ala Leu Val Glu Ile Thr Pro Pro Ile Ser Cys Gly
140                 145                 150                 155

cgc ttc att ggg ccg ggc tgc ctg ccc cac ttt aag gca ggc ctc ccc         530
Arg Phe Ile Gly Pro Gly Cys Leu Pro His Phe Lys Ala Gly Leu Pro
                160                 165                 170

aga ggc tcc cag agc tgc tgg gtg gcc ggc tgg gga tat ata gaa gag         578
Arg Gly Ser Gln Ser Cys Trp Val Ala Gly Trp Gly Tyr Ile Glu Glu
            175                 180                 185

aaa gcc ccc agg cca tca tct ata ctg atg gag gca cgt gtg gat ctc         626
Lys Ala Pro Arg Pro Ser Ser Ile Leu Met Glu Ala Arg Val Asp Leu
        190                 195                 200

atc gac ctg gac ttg tgt aac tcg acc cag tgg tac aat ggg cgc gtt         674
Ile Asp Leu Asp Leu Cys Asn Ser Thr Gln Trp Tyr Asn Gly Arg Val
    205                 210                 215

cag cca acc aat gtg tgc gcg ggg tat cct gta ggc aag atc gac acc         722
Gln Pro Thr Asn Val Cys Ala Gly Tyr Pro Val Gly Lys Ile Asp Thr
220                 225                 230                 235
```

```
tgc cag gga gac agc ggc ggg cct ctc atg tgc aaa gac agc aag gaa      770
Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Lys Asp Ser Lys Glu
                240                 245                 250

agc gcc tat gtg gtc gtg gga atc aca agc tgg ggg gta ggc tgt gcc      818
Ser Ala Tyr Val Val Val Gly Ile Thr Ser Trp Gly Val Gly Cys Ala
            255                 260                 265

cgt gcc aag cgc ccc gga atc tac acg gcc acc tgg ccc tat ctg aac      866
Arg Ala Lys Arg Pro Gly Ile Tyr Thr Ala Thr Trp Pro Tyr Leu Asn
        270                 275                 280

tgg atc gcc tcc aag att ggt tct aac gct ttg cgt atg att caa tcg      914
Trp Ile Ala Ser Lys Ile Gly Ser Asn Ala Leu Arg Met Ile Gln Ser
    285                 290                 295

gcc acc cct cca cct ccc acc act cga ccg ccc ccg att cga ccc ccc      962
Ala Thr Pro Pro Pro Pro Thr Thr Arg Pro Pro Pro Ile Arg Pro Pro
300                 305                 310                 315

ttc tcc cac cct atc tct gct cac ctt cct tgg tat ttc caa ccg ccc     1010
Phe Ser His Pro Ile Ser Ala His Leu Pro Trp Tyr Phe Gln Pro Pro
                320                 325                 330

cct cga cca ctt cca ccc cga cca ccg gca gcc cag ccc cga ccc cca     1058
Pro Arg Pro Leu Pro Pro Arg Pro Pro Ala Ala Gln Pro Arg Pro Pro
            335                 340                 345

cct tca ccc ccg ccc cca ccc cca cct cca gcc tca cct tta ccc cca     1106
Pro Ser Pro Pro Pro Pro Pro Pro Pro Pro Ala Ser Pro Leu Pro Pro
        350                 355                 360

ccc cca ccc cca ccc cca cct aca ccc tca tct acc aca aaa ctt ccc     1154
Pro Pro Pro Pro Pro Pro Pro Thr Pro Ser Ser Thr Thr Lys Leu Pro
    365                 370                 375

caa gga ctt tct ttt gcc aag cgc cta cag cag ctc ata gag gtc ttg     1202
Gln Gly Leu Ser Phe Ala Lys Arg Leu Gln Gln Leu Ile Glu Val Leu
380                 385                 390                 395

aag ggg aag acc tat tcc gac gga aag aac cat tat gac atg gag acc     1250
Lys Gly Lys Thr Tyr Ser Asp Gly Lys Asn His Tyr Asp Met Glu Thr
                400                 405                 410

aca gag ctc cca gaa ctg acc tcg acc tcc tga tctgacctgg ttctcaacag   1303
Thr Glu Leu Pro Glu Leu Thr Ser Thr Ser
            415                 420

acccagtgag cccttcactc ctgagaaaaa ggaaagatga aataaataaa taaacatata   1363

tatatagata ta                                                       1375
```

<210> 6
<211> 421
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Val Glu Met Leu Pro Thr Ala Ile Leu Leu Val Leu Ala Val Ser
1               5                  10                  15

Val Val Ala Lys Asp Asn Ala Thr Cys Asp Gly Pro Cys Gly Leu Arg
            20                  25                  30

Phe Arg Gln Asn Pro Gln Gly Gly Val Arg Ile Val Gly Gly Lys Ala
        35                  40                  45

Ala Gln His Gly Ala Trp Pro Trp Met Val Ser Leu Gln Ile Phe Thr
    50                  55                  60

Tyr Asn Ser His Arg Tyr His Thr Cys Gly Gly Ser Leu Leu Asn Ser
65                  70                  75                  80

Arg Trp Val Leu Thr Ala Ala His Cys Phe Val Gly Lys Asn Asn Val
                85                  90                  95

His Asp Trp Arg Leu Val Phe Gly Ala Lys Glu Ile Thr Tyr Gly Asn
            100                 105                 110

Asn Lys Pro Val Lys Ala Pro Leu Gln Glu Arg Tyr Val Glu Lys Ile
        115                 120                 125

Ile Ile His Glu Lys Tyr Asn Ser Ala Thr Glu Gly Asn Asp Ile Ala
    130                 135                 140

Leu Val Glu Ile Thr Pro Pro Ile Ser Cys Gly Arg Phe Ile Gly Pro
145                 150                 155                 160

Gly Cys Leu Pro His Phe Lys Ala Gly Leu Pro Arg Gly Ser Gln Ser
                165                 170                 175

Cys Trp Val Ala Gly Trp Gly Tyr Ile Glu Glu Lys Ala Pro Arg Pro
            180                 185                 190

Ser Ser Ile Leu Met Glu Ala Arg Val Asp Leu Ile Asp Leu Asp Leu
        195                 200                 205

Cys Asn Ser Thr Gln Trp Tyr Asn Gly Arg Val Gln Pro Thr Asn Val
    210                 215                 220

Cys Ala Gly Tyr Pro Val Gly Lys Ile Asp Thr Cys Gln Gly Asp Ser
225                 230                 235                 240
```

Gly Gly Pro Leu Met Cys Lys Asp Ser Lys Glu Ser Ala Tyr Val Val
245 250 255

Val Gly Ile Thr Ser Trp Gly Val Gly Cys Ala Arg Ala Lys Arg Pro
260 265 270

Gly Ile Tyr Thr Ala Thr Trp Pro Tyr Leu Asn Trp Ile Ala Ser Lys
275 280 285

Ile Gly Ser Asn Ala Leu Arg Met Ile Gln Ser Ala Thr Pro Pro Pro
290 295 300

Pro Thr Thr Arg Pro Pro Pro Ile Arg Pro Pro Phe Ser His Pro Ile
305 310 315 320

Ser Ala His Leu Pro Trp Tyr Phe Gln Pro Pro Pro Arg Pro Leu Pro
325 330 335

Pro Arg Pro Pro Ala Ala Gln Pro Arg Pro Pro Pro Ser Pro Pro Pro
340 345 350

Pro Pro Pro Pro Pro Ala Ser Pro Leu Pro Pro Pro Pro Pro Pro Pro
355 360 365

Pro Pro Thr Pro Ser Ser Thr Thr Lys Leu Pro Gln Gly Leu Ser Phe
370 375 380

Ala Lys Arg Leu Gln Gln Leu Ile Glu Val Leu Lys Gly Lys Thr Tyr
385 390 395 400

Ser Asp Gly Lys Asn His Tyr Asp Met Glu Thr Thr Glu Leu Pro Glu
405 410 415

Leu Thr Ser Thr Ser
420

<210> 7
<211> 3382
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (110)..(2035)
<223>

<400> 7

```
aggtgggaag gcagttatga cagttgagaa gtagtagaag acacggaagg cacagaaggc      60

agacttcgct cagcacaaag aagaatttc tgataaccat actggcaaa atg aac tgg      118
                                                    Met Asn Trp
                                                    1

gct gcc ttc gga ggg tct gaa ttc ttc atc cca gaa ggc att cag ata      166
Ala Ala Phe Gly Gly Ser Glu Phe Phe Ile Pro Glu Gly Ile Gln Ile
    5                  10                  15

gat tcg aga tgc cca cta agc aga aat atc acg gaa tgg tac cat tac      214
Asp Ser Arg Cys Pro Leu Ser Arg Asn Ile Thr Glu Trp Tyr His Tyr
20                 25                  30                  35

tat ggc att gca gtg aaa gaa agc tcc caa tat tat gat gtg atg tat      262
Tyr Gly Ile Ala Val Lys Glu Ser Ser Gln Tyr Tyr Asp Val Met Tyr
                40                  45                  50

tcc aag aaa gga gct gcc tgg gtg agt gag acg ggc aag aaa gaa gtg      310
Ser Lys Lys Gly Ala Ala Trp Val Ser Glu Thr Gly Lys Lys Glu Val
            55                  60                  65

agc aaa cag aca gtg gca tat tca ccc cgg tcc aaa ctc gga aca ctg      358
Ser Lys Gln Thr Val Ala Tyr Ser Pro Arg Ser Lys Leu Gly Thr Leu
        70                  75                  80

ctg cca gaa ttt ccc aat gtc aaa gat cta aat ctg cca gcc agc ctg      406
Leu Pro Glu Phe Pro Asn Val Lys Asp Leu Asn Leu Pro Ala Ser Leu
    85                  90                  95

cct gag gag aag gtt tct acc ttt att atg atg tac aga aca cac tgt      454
Pro Glu Glu Lys Val Ser Thr Phe Ile Met Met Tyr Arg Thr His Cys
100                 105                 110                 115

cag aga ata ctg gac act gta ata aga gcc aac ttt gat gag gtt caa      502
Gln Arg Ile Leu Asp Thr Val Ile Arg Ala Asn Phe Asp Glu Val Gln
                120                 125                 130

agt ttc ctt ctg cac ttt tgg caa gga atg ccg ccc cac atg ctg cct      550
Ser Phe Leu Leu His Phe Trp Gln Gly Met Pro Pro His Met Leu Pro
            135                 140                 145

gtg ctg ggc tcc tcc acg gtg gtg aac att gtc ggc gtg tgt gac tcc      598
Val Leu Gly Ser Ser Thr Val Val Asn Ile Val Gly Val Cys Asp Ser
        150                 155                 160

atc ctc tac aaa gct atc tcc ggg gtg ctg atg ccc act gtg ctg cag      646
Ile Leu Tyr Lys Ala Ile Ser Gly Val Leu Met Pro Thr Val Leu Gln
    165                 170                 175

gca tta cct gac agc tta act cag gtg att cga aag ttt gcc aag caa      694
Ala Leu Pro Asp Ser Leu Thr Gln Val Ile Arg Lys Phe Ala Lys Gln
180                 185                 190                 195

ctg gat gag tgg cta aaa gtg gct ctc cac gac ctc cca gaa aac ttg      742
Leu Asp Glu Trp Leu Lys Val Ala Leu His Asp Leu Pro Glu Asn Leu
                200                 205                 210

cga aac atc aag ttc gaa ttg tcg aga agg ttc tcc caa att ctg aga      790
Arg Asn Ile Lys Phe Glu Leu Ser Arg Arg Phe Ser Gln Ile Leu Arg
```

```
            215                 220                 225

cgg caa aca tca cta aat cat ctc tgc cag gca tct cga aca gtg atc      838
Arg Gln Thr Ser Leu Asn His Leu Cys Gln Ala Ser Arg Thr Val Ile
        230                 235                 240

cac agt gca gac atc acg ttc caa atg ctg gaa gac tgg agg aac gtg      886
His Ser Ala Asp Ile Thr Phe Gln Met Leu Glu Asp Trp Arg Asn Val
    245                 250                 255

gac ctg aac agc atc acc aag caa acc ctt tac acc atg gaa gac tct      934
Asp Leu Asn Ser Ile Thr Lys Gln Thr Leu Tyr Thr Met Glu Asp Ser
260                 265                 270                 275

cgc gat gag cac cgg aaa ctc atc acc caa tta tat cag gag ttt gac      982
Arg Asp Glu His Arg Lys Leu Ile Thr Gln Leu Tyr Gln Glu Phe Asp
                280                 285                 290

cat ctc ttg gag gag cag tct ccc atc gag tcc tac att gag tgg ctg     1030
His Leu Leu Glu Glu Gln Ser Pro Ile Glu Ser Tyr Ile Glu Trp Leu
            295                 300                 305

gat acc atg gtt gac cgc tgt gtt gtg aag gtg gct gcc aag aga cga     1078
Asp Thr Met Val Asp Arg Cys Val Val Lys Val Ala Ala Lys Arg Arg
        310                 315                 320

ggg tcc ttg aag aaa gtg gcc cag cag ttc ctc ttg atg tgg tcc tgt     1126
Gly Ser Leu Lys Lys Val Ala Gln Gln Phe Leu Leu Met Trp Ser Cys
    325                 330                 335

ttc ggc aca agg gtg atc cgg gac atg acc ttg cac agc gcc ccc agc     1174
Phe Gly Thr Arg Val Ile Arg Asp Met Thr Leu His Ser Ala Pro Ser
340                 345                 350                 355

ttc ggg tct ttt cac cta att cac tta atg ttt gat gac tac gtg ctc     1222
Phe Gly Ser Phe His Leu Ile His Leu Met Phe Asp Asp Tyr Val Leu
                360                 365                 370

tac ctg tta gaa tct ctg cac tgt cag gag cgg gcc aat gag ctc atg     1270
Tyr Leu Leu Glu Ser Leu His Cys Gln Glu Arg Ala Asn Glu Leu Met
            375                 380                 385

cga gcc atg aag gga gaa gga agc act gca gaa gtc cga gaa gag atc     1318
Arg Ala Met Lys Gly Glu Gly Ser Thr Ala Glu Val Arg Glu Glu Ile
        390                 395                 400

atc ttg aca gag gct gcc gca cca acc cct tca cca gtg cca tcg ttt     1366
Ile Leu Thr Glu Ala Ala Ala Pro Thr Pro Ser Pro Val Pro Ser Phe
    405                 410                 415

tct cca gca aaa tct gcc aca tct gtg gaa gtg cca cct ccc tct tcc     1414
Ser Pro Ala Lys Ser Ala Thr Ser Val Glu Val Pro Pro Pro Ser Ser
420                 425                 430                 435

cct gtt agc aat cct tcc cct gag tac act ggc ctc agc act aca gga     1462
Pro Val Ser Asn Pro Ser Pro Glu Tyr Thr Gly Leu Ser Thr Thr Gly
                440                 445                 450

gca atg cag gct tac acg tgg tct cta aca tac aca gtg acg acg gct     1510
Ala Met Gln Ala Tyr Thr Trp Ser Leu Thr Tyr Thr Val Thr Thr Ala
```

```
            455                 460                 465

gct ggg tcc cca gct gag aac tcc caa cag ctg ccc tgt atg agg aac     1558
Ala Gly Ser Pro Ala Glu Asn Ser Gln Gln Leu Pro Cys Met Arg Asn
        470                 475                 480

act cac gtg cct tct tcc tcc gtc aca cac agg ata cca gtt tat ccc     1606
Thr His Val Pro Ser Ser Ser Val Thr His Arg Ile Pro Val Tyr Pro
    485                 490                 495

cac aga gag gaa cat gga tac acg gga agc tat aac tat ggg agc tat     1654
His Arg Glu Glu His Gly Tyr Thr Gly Ser Tyr Asn Tyr Gly Ser Tyr
500                 505                 510                 515

ggc aac cag cat cct cac ccc atg cag agc cag tat ccg gcc ctc cct     1702
Gly Asn Gln His Pro His Pro Met Gln Ser Gln Tyr Pro Ala Leu Pro
                520                 525                 530

cat gac aca gct atc tct ggg cca ctc cac tat gcc cct tac cac agg     1750
His Asp Thr Ala Ile Ser Gly Pro Leu His Tyr Ala Pro Tyr His Arg
            535                 540                 545

agc tct gca cag tac cct ttt aat agc ccc act tcc cgg atg gaa cct     1798
Ser Ser Ala Gln Tyr Pro Phe Asn Ser Pro Thr Ser Arg Met Glu Pro
        550                 555                 560

tgt ttg atg agc agt act ccc aga ctg cat cct acc cca gtc act ccc     1846
Cys Leu Met Ser Ser Thr Pro Arg Leu His Pro Thr Pro Val Thr Pro
    565                 570                 575

cgc tgg cca gag gtg ccc tca gcc aac acg tgc tac aca aac ccg tct     1894
Arg Trp Pro Glu Val Pro Ser Ala Asn Thr Cys Tyr Thr Asn Pro Ser
580                 585                 590                 595

gtg cat tct gcg agg tac gga aac tct agt gac atg tat aca cct ctg     1942
Val His Ser Ala Arg Tyr Gly Asn Ser Ser Asp Met Tyr Thr Pro Leu
                600                 605                 610

aca acg cgc agg aat tct gaa tat gag cac atg caa cac ttt cct ggc     1990
Thr Thr Arg Arg Asn Ser Glu Tyr Glu His Met Gln His Phe Pro Gly
            615                 620                 625

ttt gct tac atc aac gga gag gcc tct aca gga tgg gct aaa tga         2035
Phe Ala Tyr Ile Asn Gly Glu Ala Ser Thr Gly Trp Ala Lys
        630                 635                 640

ctgctatcat aggcatccat atttaatatt aataataata attaataata ataataaacc   2095

caacacccat cccccagaag actttatctc tatacattgt aactcatggg ctattcctaa   2155

gtgcccattt tcctaatgaa catgaggatg ggatcaatgt gggatgaata aactttagtt   2215

cagaaacagg acttactaaa agtcagtggg actgggtttc tgtagccaag ccagacttga   2275

ctgtttctgt agagcactat ctcgggcagg ccattctgtg ccttttccct ctgttccatg   2335

actttgcttt gtgttggcaa ccacttctag taagctactg attttcctgt tgacaaaatc   2395

tctttagtct tgaaggatgg atactggaga cagaatctgg tttgtgttct tggatgggca   2455
```

cataatttac caagagcatt caccttgcca tctgtcttgt cattgtactg tacaaggaac 2515

agccctcaga cgtgttctgc acatcccttc ttcctggtgg taccatccct atttcctgga 2575

gcaccagggc taaatgggga gctatctgga aactctagat tttctgtcat acccacatct 2635

gtcacagtac ctgcattgtc ttggaatgta agcactgtct tgagggaagg aagaggtctg 2695

ttctgtattg ccttaagttg attgaggttt gtaggagact ggttcttcta catacaagga 2755

tttgtcttaa gtttgcacaa tggctagtgt cagcaaaagg caggagaggg tttttgtttt 2815

ttttttaagt tctatgagaa tgtggattta tggcattgag tatcacactc agctctgctg 2875

tgttaacttt gtgaaactgg atggaacaaa ctttaactta ccaagcacca agtgtgaaag 2935

tgactttcac ggttccttca taaaactata ataatatccg acactttgat agaaaaaaat 2995

tcaaagctgt gcctttgagc ctatactata ctgtgtatgt gtggaaataa aaatgtattg 3055

tacttttgga gaatttttg taggcatttt tctgtcagat ttgtagtaat ttgtgaggtt 3115

tgttagagat taatataggt tttctttctg tattataaaa tgcaccaagc aattatggtg 3175

gacctattac cctatgggta agaaataaat ggaaatatga catcggatgt ttcagcaact 3235

gttctgtaaa taaaatcttt gatcacacca ctcagtgtga taattgtgtc tacagctaaa 3295

atggaaatag ttttatctgt acagttgtgc aagatatgaa tggtttcaca ctcaaataaa 3355

aaatattgaa cccccaaaaa aaaaaaa 3382

<210> 8
<211> 641
<212> PRT
<213> Homo sapiens

<400> 8

Met Asn Trp Ala Ala Phe Gly Gly Ser Glu Phe Phe Ile Pro Glu Gly
1 5 10 15

Ile Gln Ile Asp Ser Arg Cys Pro Leu Ser Arg Asn Ile Thr Glu Trp
20 25 30

Tyr His Tyr Tyr Gly Ile Ala Val Lys Glu Ser Ser Gln Tyr Tyr Asp
35 40 45

Val Met Tyr Ser Lys Lys Gly Ala Ala Trp Val Ser Glu Thr Gly Lys
50 55 60

Lys Glu Val Ser Lys Gln Thr Val Ala Tyr Ser Pro Arg Ser Lys Leu
65 70 75 80

```
Gly Thr Leu Leu Pro Glu Phe Pro Asn Val Lys Asp Leu Asn Leu Pro
                85                  90                  95

Ala Ser Leu Pro Glu Glu Lys Val Ser Thr Phe Ile Met Met Tyr Arg
            100                 105                 110

Thr His Cys Gln Arg Ile Leu Asp Thr Val Ile Arg Ala Asn Phe Asp
        115                 120                 125

Glu Val Gln Ser Phe Leu Leu His Phe Trp Gln Gly Met Pro Pro His
    130                 135                 140

Met Leu Pro Val Leu Gly Ser Ser Thr Val Val Asn Ile Val Gly Val
145                 150                 155                 160

Cys Asp Ser Ile Leu Tyr Lys Ala Ile Ser Gly Val Leu Met Pro Thr
                165                 170                 175

Val Leu Gln Ala Leu Pro Asp Ser Leu Thr Gln Val Ile Arg Lys Phe
            180                 185                 190

Ala Lys Gln Leu Asp Glu Trp Leu Lys Val Ala Leu His Asp Leu Pro
        195                 200                 205

Glu Asn Leu Arg Asn Ile Lys Phe Glu Leu Ser Arg Arg Phe Ser Gln
    210                 215                 220

Ile Leu Arg Arg Gln Thr Ser Leu Asn His Leu Cys Gln Ala Ser Arg
225                 230                 235                 240

Thr Val Ile His Ser Ala Asp Ile Thr Phe Gln Met Leu Glu Asp Trp
                245                 250                 255

Arg Asn Val Asp Leu Asn Ser Ile Thr Lys Gln Thr Leu Tyr Thr Met
            260                 265                 270

Glu Asp Ser Arg Asp Glu His Arg Lys Leu Ile Thr Gln Leu Tyr Gln
        275                 280                 285

Glu Phe Asp His Leu Leu Glu Glu Gln Ser Pro Ile Glu Ser Tyr Ile
    290                 295                 300

Glu Trp Leu Asp Thr Met Val Asp Arg Cys Val Val Lys Val Ala Ala
305                 310                 315                 320
```

Lys Arg Arg Gly Ser Leu Lys Lys Val Ala Gln Gln Phe Leu Leu Met
325 330 335

Trp Ser Cys Phe Gly Thr Arg Val Ile Arg Asp Met Thr Leu His Ser
340 345 350

Ala Pro Ser Phe Gly Ser Phe His Leu Ile His Leu Met Phe Asp Asp
355 360 365

Tyr Val Leu Tyr Leu Leu Glu Ser Leu His Cys Gln Glu Arg Ala Asn
370 375 380

Glu Leu Met Arg Ala Met Lys Gly Glu Gly Ser Thr Ala Glu Val Arg
385 390 395 400

Glu Glu Ile Ile Leu Thr Glu Ala Ala Ala Pro Thr Pro Ser Pro Val
405 410 415

Pro Ser Phe Ser Pro Ala Lys Ser Ala Thr Ser Val Glu Val Pro Pro
420 425 430

Pro Ser Ser Pro Val Ser Asn Pro Ser Pro Glu Tyr Thr Gly Leu Ser
435 440 445

Thr Thr Gly Ala Met Gln Ala Tyr Thr Trp Ser Leu Thr Tyr Thr Val
450 455 460

Thr Thr Ala Ala Gly Ser Pro Ala Glu Asn Ser Gln Gln Leu Pro Cys
465 470 475 480

Met Arg Asn Thr His Val Pro Ser Ser Ser Val Thr His Arg Ile Pro
485 490 495

Val Tyr Pro His Arg Glu Glu His Gly Tyr Thr Gly Ser Tyr Asn Tyr
500 505 510

Gly Ser Tyr Gly Asn Gln His Pro His Pro Met Gln Ser Gln Tyr Pro
515 520 525

Ala Leu Pro His Asp Thr Ala Ile Ser Gly Pro Leu His Tyr Ala Pro
530 535 540

Tyr His Arg Ser Ser Ala Gln Tyr Pro Phe Asn Ser Pro Thr Ser Arg
545 550 555 560

```
Met Glu Pro Cys Leu Met Ser Ser Thr Pro Arg Leu His Pro Thr Pro
                565                 570                 575

Val Thr Pro Arg Trp Pro Glu Val Pro Ser Ala Asn Thr Cys Tyr Thr
            580                 585                 590

Asn Pro Ser Val His Ser Ala Arg Tyr Gly Asn Ser Ser Asp Met Tyr
        595                 600                 605

Thr Pro Leu Thr Thr Arg Arg Asn Ser Glu Tyr Glu His Met Gln His
    610                 615                 620

Phe Pro Gly Phe Ala Tyr Ile Asn Gly Glu Ala Ser Thr Gly Trp Ala
625                 630                 635                 640

Lys
```

<210> 9
<211> 2186
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (106)..(1797)
<223>

<400> 9

```
tggagaggcc acagctgctg gcttcctggg cttctccaaa ctcctgtgtg tcgccactgc     60

caccggcagg gagccaggag agagacagaa aggggctgag acaga atg atc aaa agg     117
                                                  Met Ile Lys Arg
                                                  1

aga gcc cac cct ggt gcg gga ggc gac agg acc agg cct cga cgg cgc       165
Arg Ala His Pro Gly Ala Gly Gly Asp Arg Thr Arg Pro Arg Arg Arg
5                   10                  15                  20

cgt tcc act gag agc tgg att gaa aga tgt ctc aac gaa agt gaa aac       213
Arg Ser Thr Glu Ser Trp Ile Glu Arg Cys Leu Asn Glu Ser Glu Asn
                25                  30                  35

aaa cgt tat tcc agc cac aca tct ctg ggg aat gtt tct aat gat gaa       261
Lys Arg Tyr Ser Ser His Thr Ser Leu Gly Asn Val Ser Asn Asp Glu
            40                  45                  50

aat gag gaa aaa gaa aat aat aga gca tcc aag ccc cac tcc act cct       309
Asn Glu Glu Lys Glu Asn Asn Arg Ala Ser Lys Pro His Ser Thr Pro
        55                  60                  65

gct act ctg caa tgg ctg gag gag aac tat gag att gca gag ggg gtc       357
Ala Thr Leu Gln Trp Leu Glu Glu Asn Tyr Glu Ile Ala Glu Gly Val
    70                  75                  80
```

```
tgc atc cct cgc agt gcc ctc tat atg cat tac ctg gat ttc tgc gag      405
Cys Ile Pro Arg Ser Ala Leu Tyr Met His Tyr Leu Asp Phe Cys Glu
85                  90                  95                  100

aag aat gat acc caa cct gtc aat gct gcc agc ttt gga aag atc ata      453
Lys Asn Asp Thr Gln Pro Val Asn Ala Ala Ser Phe Gly Lys Ile Ile
                105                 110                 115

agg cag cag ttt cct cag tta acc acc aga aga ctc ggg acc cga gga      501
Arg Gln Gln Phe Pro Gln Leu Thr Thr Arg Arg Leu Gly Thr Arg Gly
            120                 125                 130

cag tca aag tac cat tac tat ggc att gca gtg aaa gaa agc tcc caa      549
Gln Ser Lys Tyr His Tyr Tyr Gly Ile Ala Val Lys Glu Ser Ser Gln
        135                 140                 145

tat tat gat gtg atg tat tcc aag aaa gga gct gcc tgg gtg agt gag      597
Tyr Tyr Asp Val Met Tyr Ser Lys Lys Gly Ala Ala Trp Val Ser Glu
    150                 155                 160

acg ggc aag aaa gaa gtg agc aaa cag aca gtg gca tat tca ccc cgg      645
Thr Gly Lys Lys Glu Val Ser Lys Gln Thr Val Ala Tyr Ser Pro Arg
165                 170                 175                 180

tcc aaa ctc gga aca ctg ctg cca gaa ttt ccc aat gtc aaa gat cta      693
Ser Lys Leu Gly Thr Leu Leu Pro Glu Phe Pro Asn Val Lys Asp Leu
                185                 190                 195

aat ctg cca gcc agc ctg cct gag gag aag gtt tct acc ttt att atg      741
Asn Leu Pro Ala Ser Leu Pro Glu Glu Lys Val Ser Thr Phe Ile Met
            200                 205                 210

atg tac aga aca cac tgt cag aga ata ctg gac act gta ata aga gcc      789
Met Tyr Arg Thr His Cys Gln Arg Ile Leu Asp Thr Val Ile Arg Ala
        215                 220                 225

aac ttt gat gag gtt caa agt ttc ctt ctg cac ttt tgg caa gga atg      837
Asn Phe Asp Glu Val Gln Ser Phe Leu Leu His Phe Trp Gln Gly Met
    230                 235                 240

ccg ccc cac atg ctg cct gtg ctg ggc tcc tcc acg gtg gtg aac att      885
Pro Pro His Met Leu Pro Val Leu Gly Ser Ser Thr Val Val Asn Ile
245                 250                 255                 260

gtc ggc gtg tgt gac tcc atc ctc tac aaa gct atc tcc ggg gtg ctg      933
Val Gly Val Cys Asp Ser Ile Leu Tyr Lys Ala Ile Ser Gly Val Leu
                265                 270                 275

atg ccc act gtg ctg cag gca tta cct gac agc tta act cag gtg att      981
Met Pro Thr Val Leu Gln Ala Leu Pro Asp Ser Leu Thr Gln Val Ile
            280                 285                 290

cga aag ttt gcc aag caa ctg gat gag tgg cta aaa gtg gct ctc cac     1029
Arg Lys Phe Ala Lys Gln Leu Asp Glu Trp Leu Lys Val Ala Leu His
        295                 300                 305

gac ctc cca gaa aac ttg cga aac atc aag ttc gaa ttg tcg aga agg     1077
Asp Leu Pro Glu Asn Leu Arg Asn Ile Lys Phe Glu Leu Ser Arg Arg
    310                 315                 320
```

```
ttc tcc caa att ctg aga cgg caa aca tca cta aat cat ctc tgc cag     1125
Phe Ser Gln Ile Leu Arg Arg Gln Thr Ser Leu Asn His Leu Cys Gln
325                 330                 335                 340

gca tct cga aca gtg atc cac agt gca gac atc acg ttc caa atg ctg     1173
Ala Ser Arg Thr Val Ile His Ser Ala Asp Ile Thr Phe Gln Met Leu
                345                 350                 355

gaa gac tgg agg aac gtg gac ctg aac agc atc acc aag caa acc ctt     1221
Glu Asp Trp Arg Asn Val Asp Leu Asn Ser Ile Thr Lys Gln Thr Leu
            360                 365                 370

tac acc atg gaa gac tct cgc gat gag cac cgg aaa ctc atc acc caa     1269
Tyr Thr Met Glu Asp Ser Arg Asp Glu His Arg Lys Leu Ile Thr Gln
        375                 380                 385

tta tat cag gag ttt gac cat ctc ttg gag gag cag tct ccc atc gag     1317
Leu Tyr Gln Glu Phe Asp His Leu Leu Glu Glu Gln Ser Pro Ile Glu
    390                 395                 400

tcc tac att gag tgg ctg gat acc atg gtt gac cgc tgt gtt gtg aag     1365
Ser Tyr Ile Glu Trp Leu Asp Thr Met Val Asp Arg Cys Val Val Lys
405                 410                 415                 420

gtg gct gcc aag aga caa ggg tcc ttg aag aaa gtg gcc cag cag ttc     1413
Val Ala Ala Lys Arg Gln Gly Ser Leu Lys Lys Val Ala Gln Gln Phe
                425                 430                 435

ctc ttg atg tgg tcc tgt ttc ggc aca agg gtg atc cgg gac atg acc     1461
Leu Leu Met Trp Ser Cys Phe Gly Thr Arg Val Ile Arg Asp Met Thr
            440                 445                 450

ttg cac agc gcc ccc agc ttc ggg tct ttt cac cta att cac tta atg     1509
Leu His Ser Ala Pro Ser Phe Gly Ser Phe His Leu Ile His Leu Met
        455                 460                 465

ttt gat gac tac gtg ctc tac ctg tta gaa tct ctg cac tgt cag gag     1557
Phe Asp Asp Tyr Val Leu Tyr Leu Leu Glu Ser Leu His Cys Gln Glu
    470                 475                 480

cgg gcc aat gag ctc atg cga gcc atg aag gga gaa gga agc act gca     1605
Arg Ala Asn Glu Leu Met Arg Ala Met Lys Gly Glu Gly Ser Thr Ala
485                 490                 495                 500

gaa gtc cga gaa gag atc atc ttg aca gag gct gcc gca cca acc cct     1653
Glu Val Arg Glu Glu Ile Ile Leu Thr Glu Ala Ala Ala Pro Thr Pro
                505                 510                 515

tca cca gtg cca tcg ttt tct cca gca aaa tct gcc aca tct gtg gaa     1701
Ser Pro Val Pro Ser Phe Ser Pro Ala Lys Ser Ala Thr Ser Val Glu
            520                 525                 530

gtg cca cct ccc tct tcc cct gtt agc aat cct tcc cct gag tac act     1749
Val Pro Pro Pro Ser Ser Pro Val Ser Asn Pro Ser Pro Glu Tyr Thr
        535                 540                 545

ggc ctc agc act aca ggt aat gga aag tcc ttc aaa aac ttt ggg tag     1797
Gly Leu Ser Thr Thr Gly Asn Gly Lys Ser Phe Lys Asn Phe Gly
    550                 555                 560
```

ttaatgtttg aagaaagggc tttctgccag cctgggcaac atagtgagac ttcatttcca 1857

cacacacaaa aagccagaca tcttggctca cacctgtagt cccagctact tgggaggctg 1917

aggtgggaga attgcttgag cccaggagct acgatcgcac cactgcattc tagccttagt 1977

gatacagtga gaccttgtct caaaaaagga aaaacagggc tttctggaaa aacattcttc 2037

tcccacaatc tccaaaagat aatgccaaaa cctgggtatc ttcctggatt tgtgaatgac 2097

gtacaggtat tcatttattc attggtacac attctgtatg ctgctgtttt caagttggca 2157

aattaagcat atgataaaat cccaaaact 2186

<210> 10
<211> 563
<212> PRT
<213> Homo sapiens

<400> 10

Met Ile Lys Arg Arg Ala His Pro Gly Ala Gly Gly Asp Arg Thr Arg
1 5 10 15

Pro Arg Arg Arg Arg Ser Thr Glu Ser Trp Ile Glu Arg Cys Leu Asn
20 25 30

Glu Ser Glu Asn Lys Arg Tyr Ser Ser His Thr Ser Leu Gly Asn Val
35 40 45

Ser Asn Asp Glu Asn Glu Glu Lys Glu Asn Asn Arg Ala Ser Lys Pro
50 55 60

His Ser Thr Pro Ala Thr Leu Gln Trp Leu Glu Glu Asn Tyr Glu Ile
65 70 75 80

Ala Glu Gly Val Cys Ile Pro Arg Ser Ala Leu Tyr Met His Tyr Leu
85 90 95

Asp Phe Cys Glu Lys Asn Asp Thr Gln Pro Val Asn Ala Ala Ser Phe
100 105 110

Gly Lys Ile Ile Arg Gln Gln Phe Pro Gln Leu Thr Thr Arg Arg Leu
115 120 125

Gly Thr Arg Gly Gln Ser Lys Tyr His Tyr Tyr Gly Ile Ala Val Lys
130 135 140

Glu Ser Ser Gln Tyr Tyr Asp Val Met Tyr Ser Lys Lys Gly Ala Ala

145 150 155 160

Trp Val Ser Glu Thr Gly Lys Lys Glu Val Ser Lys Gln Thr Val Ala
165 170 175

Tyr Ser Pro Arg Ser Lys Leu Gly Thr Leu Leu Pro Glu Phe Pro Asn
180 185 190

Val Lys Asp Leu Asn Leu Pro Ala Ser Leu Pro Glu Glu Lys Val Ser
195 200 205

Thr Phe Ile Met Met Tyr Arg Thr His Cys Gln Arg Ile Leu Asp Thr
210 215 220

Val Ile Arg Ala Asn Phe Asp Glu Val Gln Ser Phe Leu Leu His Phe
225 230 235 240

Trp Gln Gly Met Pro Pro His Met Leu Pro Val Leu Gly Ser Ser Thr
245 250 255

Val Val Asn Ile Val Gly Val Cys Asp Ser Ile Leu Tyr Lys Ala Ile
260 265 270

Ser Gly Val Leu Met Pro Thr Val Leu Gln Ala Leu Pro Asp Ser Leu
275 280 285

Thr Gln Val Ile Arg Lys Phe Ala Lys Gln Leu Asp Glu Trp Leu Lys
290 295 300

Val Ala Leu His Asp Leu Pro Glu Asn Leu Arg Asn Ile Lys Phe Glu
305 310 315 320

Leu Ser Arg Arg Phe Ser Gln Ile Leu Arg Arg Gln Thr Ser Leu Asn
325 330 335

His Leu Cys Gln Ala Ser Arg Thr Val Ile His Ser Ala Asp Ile Thr
340 345 350

Phe Gln Met Leu Glu Asp Trp Arg Asn Val Asp Leu Asn Ser Ile Thr
355 360 365

Lys Gln Thr Leu Tyr Thr Met Glu Asp Ser Arg Asp Glu His Arg Lys
370 375 380

Leu Ile Thr Gln Leu Tyr Gln Glu Phe Asp His Leu Leu Glu Glu Gln

385 390 395 400

Ser Pro Ile Glu Ser Tyr Ile Glu Trp Leu Asp Thr Met Val Asp Arg
405 410 415

Cys Val Val Lys Val Ala Ala Lys Arg Gln Gly Ser Leu Lys Lys Val
420 425 430

Ala Gln Gln Phe Leu Leu Met Trp Ser Cys Phe Gly Thr Arg Val Ile
435 440 445

Arg Asp Met Thr Leu His Ser Ala Pro Ser Phe Gly Ser Phe His Leu
450 455 460

Ile His Leu Met Phe Asp Asp Tyr Val Leu Tyr Leu Leu Glu Ser Leu
465 470 475 480

His Cys Gln Glu Arg Ala Asn Glu Leu Met Arg Ala Met Lys Gly Glu
485 490 495

Gly Ser Thr Ala Glu Val Arg Glu Glu Ile Ile Leu Thr Glu Ala Ala
500 505 510

Ala Pro Thr Pro Ser Pro Val Pro Ser Phe Ser Pro Ala Lys Ser Ala
515 520 525

Thr Ser Val Glu Val Pro Pro Pro Ser Ser Pro Val Ser Asn Pro Ser
530 535 540

Pro Glu Tyr Thr Gly Leu Ser Thr Thr Gly Asn Gly Lys Ser Phe Lys
545 550 555 560

Asn Phe Gly

<210> 11
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 11
ccagaggaac atcaagtcag c 21

<210> 12
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 12
atattgtgcc tgtagatgtg 20

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 13
tgccgaaaat gctgaaggag 20

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 14
gtagacaaac tggaaggtgc 20

<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 15
tacattgagt ggctggatac 20

<210> 16
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 16
aggtagagca cgtagtcatc 20

<210> 17
<211> 31
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 17
cacacaggat ccatggatgc tgcagatgcg g 31

<210> 18
<211> 32
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 18
cacacaaagc ttggcttagc gcctctgccc tg 32

<210> 19
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 19
ccagaggaac atcaagtcag c 21

<210> 20
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 20
gagaaagagt tggagcaggg aa 22

<210> 21
<211> 21
<212> PRT

<213> Artificial Sequence

<220>

<223> Primer

<400> 21

Gly Gly Cys Ala Gly Thr Thr Cys Thr Thr Ala Cys Cys Ala Ala Gly
1 5 10 15

Ala Ala Gly Ala Thr
20

<210> 22
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 22
ggaggtaaaa ccagtgtcct c 21

<210> 23
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 23
tgcatgactg gagactggtt 20

<210> 24
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 24
cagttcagat aaggccaggt 20

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 25
agaggccact gagaaagcaa 20

<210> 26
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 26
ggctgctagt gtgacgttga 20

<210> 27
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 27
aaggacaggg gactaaggag 20

<210> 28
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 28
ccgtacaaat ccagcccgta 20

<210> 29
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 29
ctaacttcgg ccttcccaga 20

<210> 30
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 30
agtggggttg ccgattacag 20

<210> 31
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 31
aagcaattca ccaaggctgc 20

<210> 32
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 32
acctatcatg ccgttcttcc 20

<210> 33
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 33
aggttctact gctctccttc 20

<210> 34
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 34
gtagagaaac tggaaggtgc 20

<210> 35
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 35
atgggaatgt gtggcagtag a 21

<210> 36
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 36
ccacttacaa tttcccgtct g 21

<210> 37
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 37
actcccacca aaggcataga 20

<210> 38
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 38
cgaatcatct ctgtccatcg 20

<210> 39
<211> 20
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 39
tgtgtgactc catcctctac 20

<210> 40
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 40
aggtagagca cgtagtcatc 20

<210> 41
<211> 1886
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (49)..(1680)
<223>

<400> 41
gttagaggcg gcttgtgtcc acgggacgcg ggcggatctt ctccggcc atg agg aag 57
Met Arg Lys
1

cca gcc gct ggc ttc ctt ccc tca ctc ctg aag gtg ctg ctc ctg cct 105
Pro Ala Ala Gly Phe Leu Pro Ser Leu Leu Lys Val Leu Leu Leu Pro
5 10 15

ctg gca cct gcc gca gcc cag gat tcg act cag gcc ccc act cca ggc 153
Leu Ala Pro Ala Ala Ala Gln Asp Ser Thr Gln Ala Pro Thr Pro Gly
20 25 30 35

agc cct ctc tct cct acc gaa tac gaa cgc ttc ttc gca ctg ctg act 201
Ser Pro Leu Ser Pro Thr Glu Tyr Glu Arg Phe Phe Ala Leu Leu Thr
40 45 50

cca acc tgg aag gca gag act acc tgc cgt ctc cgt gca acc cac ggc 249
Pro Thr Trp Lys Ala Glu Thr Thr Cys Arg Leu Arg Ala Thr His Gly
55 60 65

tgc cgg aat ccc aca ctc gtc cag ctg gac caa tat gaa aac cac ggc 297
Cys Arg Asn Pro Thr Leu Val Gln Leu Asp Gln Tyr Glu Asn His Gly
70 75 80

tta gtg ccc gat ggt gct gtc tgc tcc aac ctc cct tat gcc tcc tgg 345
Leu Val Pro Asp Gly Ala Val Cys Ser Asn Leu Pro Tyr Ala Ser Trp

```
    85                  90                  95

ttt gag tct ttc tgc cag ttc act cac tac cgt tgc tcc aac cac gtc      393
Phe Glu Ser Phe Cys Gln Phe Thr His Tyr Arg Cys Ser Asn His Val
100                 105                 110                 115

tac tat gcc aag aga gtc ctg tgt tcc cag cca gtc tct att ctc tca      441
Tyr Tyr Ala Lys Arg Val Leu Cys Ser Gln Pro Val Ser Ile Leu Ser
                120                 125                 130

cct aac act ctc aag gag ata gaa gct tca gct gaa gtc tca ccc acc      489
Pro Asn Thr Leu Lys Glu Ile Glu Ala Ser Ala Glu Val Ser Pro Thr
            135                 140                 145

acg atg acc tcc ccc atc tca ccc cac ttc aca gtg aca gaa cgc cag      537
Thr Met Thr Ser Pro Ile Ser Pro His Phe Thr Val Thr Glu Arg Gln
        150                 155                 160

acc ttc cag ccc tgg cct gag agg ctc agc aac aac gtg gaa gag ctc      585
Thr Phe Gln Pro Trp Pro Glu Arg Leu Ser Asn Asn Val Glu Glu Leu
    165                 170                 175

cta caa tcc tcc ttg tcc ctg gga ggc cag gag caa gcg cca gag cac      633
Leu Gln Ser Ser Leu Ser Leu Gly Gly Gln Glu Gln Ala Pro Glu His
180                 185                 190                 195

aag cag gag caa gga gtg gag cac agg cag gag ccg aca caa gaa cac      681
Lys Gln Glu Gln Gly Val Glu His Arg Gln Glu Pro Thr Gln Glu His
                200                 205                 210

aag cag gaa gag ggg cag aaa cag gaa gag caa gaa gag gaa cag gaa      729
Lys Gln Glu Glu Gly Gln Lys Gln Glu Glu Gln Glu Glu Glu Gln Glu
            215                 220                 225

gag gag gga aag cag gaa gaa gga cag ggg act aag gag gga cgg gag      777
Glu Glu Gly Lys Gln Glu Glu Gly Gln Gly Thr Lys Glu Gly Arg Glu
        230                 235                 240

gct gtg tct cag ctg cag aca gac tca gag ccc aag ttt cac tct gaa      825
Ala Val Ser Gln Leu Gln Thr Asp Ser Glu Pro Lys Phe His Ser Glu
    245                 250                 255

tct cta tct tct aac cct tcc tct ttt gct ccc cgg gta cga gaa gta      873
Ser Leu Ser Ser Asn Pro Ser Ser Phe Ala Pro Arg Val Arg Glu Val
260                 265                 270                 275

gag tct act cct atg ata atg gag aac atc cag gag ctc att cga tca      921
Glu Ser Thr Pro Met Ile Met Glu Asn Ile Gln Glu Leu Ile Arg Ser
                280                 285                 290

gcc cag gaa ata gat gaa atg aat gaa ata tat gat gag aac tcc tac      969
Ala Gln Glu Ile Asp Glu Met Asn Glu Ile Tyr Asp Glu Asn Ser Tyr
            295                 300                 305

tgg aga aac caa aac cct ggc agc ttc ctg cag ctg ccc cac aca gag     1017
Trp Arg Asn Gln Asn Pro Gly Ser Phe Leu Gln Leu Pro His Thr Glu
        310                 315                 320

gcc ttg ctg gtg ctg tgc tat tcg atc gtg gag aat acc tgc atc ata     1065
Ala Leu Leu Val Leu Cys Tyr Ser Ile Val Glu Asn Thr Cys Ile Ile
```

```
    325                 330                 335

acc ccc aca gcc aag gcc tgg aag tac atg gag gag gag atc ctt ggt      1113
Thr Pro Thr Ala Lys Ala Trp Lys Tyr Met Glu Glu Glu Ile Leu Gly
340                 345                 350                 355

ttc ggg aag tcg gtc tgt gac agc ctt ggg cgg cga cac atg tct acc      1161
Phe Gly Lys Ser Val Cys Asp Ser Leu Gly Arg Arg His Met Ser Thr
                360                 365                 370

tgt gcc ctc tgt gac ttc tgc tcc ttg aag ctg gag cag tgc cac tca      1209
Cys Ala Leu Cys Asp Phe Cys Ser Leu Lys Leu Glu Gln Cys His Ser
            375                 380                 385

gag gcc agc ctg cag cgg caa caa tgc gac acc tcc cac aag act ccc      1257
Glu Ala Ser Leu Gln Arg Gln Gln Cys Asp Thr Ser His Lys Thr Pro
        390                 395                 400

ttt gtc agc ccc ttg ctt gcc tcc cag agc ctg tcc atc ggc aac cag      1305
Phe Val Ser Pro Leu Leu Ala Ser Gln Ser Leu Ser Ile Gly Asn Gln
    405                 410                 415

gta ggg tcc cca gaa tca ggc cgc ttt tac ggg ctg gat ttg tac ggt      1353
Val Gly Ser Pro Glu Ser Gly Arg Phe Tyr Gly Leu Asp Leu Tyr Gly
420                 425                 430                 435

ggg ctc cac atg gac ttc tgg tgt gcc cgg ctt gcc acg aaa ggc tgt      1401
Gly Leu His Met Asp Phe Trp Cys Ala Arg Leu Ala Thr Lys Gly Cys
                440                 445                 450

gaa gat gtc cga gtc tct ggg tgg ctc cag act gag ttc ctt agc ttc      1449
Glu Asp Val Arg Val Ser Gly Trp Leu Gln Thr Glu Phe Leu Ser Phe
            455                 460                 465

cag gat ggg gat ttc cct acc aag att tgt gac aca gac tat atc cag      1497
Gln Asp Gly Asp Phe Pro Thr Lys Ile Cys Asp Thr Asp Tyr Ile Gln
        470                 475                 480

tac cca aac tac tgt tcc ttc aaa agc cag cag tgt ctg atg aga aac      1545
Tyr Pro Asn Tyr Cys Ser Phe Lys Ser Gln Gln Cys Leu Met Arg Asn
    485                 490                 495

cgc aat cgg aag gtg tcc cgc atg aga tgt ctg cag aat gag act tac      1593
Arg Asn Arg Lys Val Ser Arg Met Arg Cys Leu Gln Asn Glu Thr Tyr
500                 505                 510                 515

agt gcg ctg agc cct ggc aaa agt gag gac gtt gtg ctt cga tgg agc      1641
Ser Ala Leu Ser Pro Gly Lys Ser Glu Asp Val Val Leu Arg Trp Ser
                520                 525                 530

cag gag ttc agc acc ttg act cta ggc cag ttc gga tga gctggcgtct      1690
Gln Glu Phe Ser Thr Leu Thr Leu Gly Gln Phe Gly
            535                 540

attctgccca caccccagcc caacctgccc acgttctcta ttgttttgag accccattgc  1750

tttcaggctg ccccttctgg gtctgttact cggccctac tcacatttcc ttgggttgga  1810

gcaacagtcc cagagagggc cacggtggga gctgcgccct ccttaaaaga tgactttaca  1870
```

taaaatgttg atcttc 1886

<210> 42
<211> 543
<212> PRT
<213> Homo sapiens

<400> 42

Met Arg Lys Pro Ala Ala Gly Phe Leu Pro Ser Leu Leu Lys Val Leu
1 5 10 15

Leu Leu Pro Leu Ala Pro Ala Ala Ala Gln Asp Ser Thr Gln Ala Pro
20 25 30

Thr Pro Gly Ser Pro Leu Ser Pro Thr Glu Tyr Glu Arg Phe Phe Ala
35 40 45

Leu Leu Thr Pro Thr Trp Lys Ala Glu Thr Thr Cys Arg Leu Arg Ala
50 55 60

Thr His Gly Cys Arg Asn Pro Thr Leu Val Gln Leu Asp Gln Tyr Glu
65 70 75 80

Asn His Gly Leu Val Pro Asp Gly Ala Val Cys Ser Asn Leu Pro Tyr
85 90 95

Ala Ser Trp Phe Glu Ser Phe Cys Gln Phe Thr His Tyr Arg Cys Ser
100 105 110

Asn His Val Tyr Tyr Ala Lys Arg Val Leu Cys Ser Gln Pro Val Ser
115 120 125

Ile Leu Ser Pro Asn Thr Leu Lys Glu Ile Glu Ala Ser Ala Glu Val
130 135 140

Ser Pro Thr Thr Met Thr Ser Pro Ile Ser Pro His Phe Thr Val Thr
145 150 155 160

Glu Arg Gln Thr Phe Gln Pro Trp Pro Glu Arg Leu Ser Asn Asn Val
165 170 175

Glu Glu Leu Leu Gln Ser Ser Leu Ser Leu Gly Gly Gln Glu Gln Ala
180 185 190

Pro Glu His Lys Gln Glu Gln Gly Val Glu His Arg Gln Glu Pro Thr
195 200 205

Gln Glu His Lys Gln Glu Glu Gly Gln Lys Gln Glu Glu Gln Glu Glu
210 215 220

Glu Gln Glu Glu Glu Gly Lys Gln Glu Glu Gly Gln Gly Thr Lys Glu
225 230 235 240

Gly Arg Glu Ala Val Ser Gln Leu Gln Thr Asp Ser Glu Pro Lys Phe
245 250 255

His Ser Glu Ser Leu Ser Ser Asn Pro Ser Ser Phe Ala Pro Arg Val
260 265 270

Arg Glu Val Glu Ser Thr Pro Met Ile Met Glu Asn Ile Gln Glu Leu
275 280 285

Ile Arg Ser Ala Gln Glu Ile Asp Glu Met Asn Glu Ile Tyr Asp Glu
290 295 300

Asn Ser Tyr Trp Arg Asn Gln Asn Pro Gly Ser Phe Leu Gln Leu Pro
305 310 315 320

His Thr Glu Ala Leu Leu Val Leu Cys Tyr Ser Ile Val Glu Asn Thr
325 330 335

Cys Ile Ile Thr Pro Thr Ala Lys Ala Trp Lys Tyr Met Glu Glu Glu
340 345 350

Ile Leu Gly Phe Gly Lys Ser Val Cys Asp Ser Leu Gly Arg Arg His
355 360 365

Met Ser Thr Cys Ala Leu Cys Asp Phe Cys Ser Leu Lys Leu Glu Gln
370 375 380

Cys His Ser Glu Ala Ser Leu Gln Arg Gln Gln Cys Asp Thr Ser His
385 390 395 400

Lys Thr Pro Phe Val Ser Pro Leu Leu Ala Ser Gln Ser Leu Ser Ile
405 410 415

Gly Asn Gln Val Gly Ser Pro Glu Ser Gly Arg Phe Tyr Gly Leu Asp
420 425 430

Leu Tyr Gly Gly Leu His Met Asp Phe Trp Cys Ala Arg Leu Ala Thr
435 440 445

```
Lys Gly Cys Glu Asp Val Arg Val Ser Gly Trp Leu Gln Thr Glu Phe
    450                 455                 460

Leu Ser Phe Gln Asp Gly Asp Phe Pro Thr Lys Ile Cys Asp Thr Asp
465                 470                 475                 480

Tyr Ile Gln Tyr Pro Asn Tyr Cys Ser Phe Lys Ser Gln Gln Cys Leu
                485                 490                 495

Met Arg Asn Arg Asn Arg Lys Val Ser Arg Met Arg Cys Leu Gln Asn
            500                 505                 510

Glu Thr Tyr Ser Ala Leu Ser Pro Gly Lys Ser Glu Asp Val Val Leu
        515                 520                 525

Arg Trp Ser Gln Glu Phe Ser Thr Leu Thr Leu Gly Gln Phe Gly
    530                 535                 540
```

<210> 43
<211> 1100
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (53)..(850)
<223>

<400> 43

```
gctatgaagc agctgtggcc cacactgggg tcccctcttt tcctaaatcc ag atg aac     58
                                                          Met Asn
                                                          1

agg ttt ctc ttg cta atg agt ctt tat ctg ctt gga tct gcc aga gga      106
Arg Phe Leu Leu Leu Met Ser Leu Tyr Leu Leu Gly Ser Ala Arg Gly
        5                   10                  15

aca tca agt cag cct aat gag ctt tct ggc tcc ata gat cat caa act      154
Thr Ser Ser Gln Pro Asn Glu Leu Ser Gly Ser Ile Asp His Gln Thr
    20                  25                  30

tca gtt cag caa ctt cca ggt gag ttc ttt tca ctt gaa aac cct tct      202
Ser Val Gln Gln Leu Pro Gly Glu Phe Phe Ser Leu Glu Asn Pro Ser
35                  40                  45                  50

gat gct gag gct tta tat gag act tct tca ggc ctg aac act tta agt      250
Asp Ala Glu Ala Leu Tyr Glu Thr Ser Ser Gly Leu Asn Thr Leu Ser
                55                  60                  65

gag cat ggt tcc agt gag cat ggt tca agc aag cac act gtg gcc gag      298
Glu His Gly Ser Ser Glu His Gly Ser Ser Lys His Thr Val Ala Glu
            70                  75                  80
```

```
cac act tct gga gaa cat gct gag agt gag cat gct tca ggt gag ccc      346
His Thr Ser Gly Glu His Ala Glu Ser Glu His Ala Ser Gly Glu Pro
        85                  90                  95

gct gcg act gaa cat gct gaa ggt gag cat act gta ggt gag cag cct      394
Ala Ala Thr Glu His Ala Glu Gly Glu His Thr Val Gly Glu Gln Pro
    100                 105                 110

tca gga gaa cag cct tca ggt gaa cac ctc tcc gga gaa cag cct ttg      442
Ser Gly Glu Gln Pro Ser Gly Glu His Leu Ser Gly Glu Gln Pro Leu
115                 120                 125                 130

agt gag ctt gag tca ggt gaa cag cct tca gat gaa cag cct tca ggt      490
Ser Glu Leu Glu Ser Gly Glu Gln Pro Ser Asp Glu Gln Pro Ser Gly
                135                 140                 145

gaa cat ggc tcc ggt gaa cag cct tct ggt gag cag gcc tcg ggt gaa      538
Glu His Gly Ser Gly Glu Gln Pro Ser Gly Glu Gln Ala Ser Gly Glu
            150                 155                 160

cag cct tca ggt gag cac gct tca ggg gaa cag gct tca ggt gca cca      586
Gln Pro Ser Gly Glu His Ala Ser Gly Glu Gln Ala Ser Gly Ala Pro
        165                 170                 175

att tca agc aca tct aca ggc aca ata tta aat tgc tac aca tgt gct      634
Ile Ser Ser Thr Ser Thr Gly Thr Ile Leu Asn Cys Tyr Thr Cys Ala
    180                 185                 190

tat atg aat gat caa gga aaa tgt ctt cgt gga gag gga acc tgc atc      682
Tyr Met Asn Asp Gln Gly Lys Cys Leu Arg Gly Glu Gly Thr Cys Ile
195                 200                 205                 210

act cag aat tcc cag cag tgc atg tta aag aag atc ttt gaa ggt gga      730
Thr Gln Asn Ser Gln Gln Cys Met Leu Lys Lys Ile Phe Glu Gly Gly
                215                 220                 225

aaa ctc caa ttc atg gtt caa ggg tgt gag aac atg tgc cca tct atg      778
Lys Leu Gln Phe Met Val Gln Gly Cys Glu Asn Met Cys Pro Ser Met
            230                 235                 240

aac ctc ttc tcc cat gga acg agg atg caa att ata tgc tgt cga aat      826
Asn Leu Phe Ser His Gly Thr Arg Met Gln Ile Ile Cys Cys Arg Asn
        245                 250                 255

caa tct ttc tgc aat aag atc tag aagcctgggc ccttgcttgt tttgactcag     880
Gln Ser Phe Cys Asn Lys Ile
    260                 265

gcagtaaaaa gcctccatca ctctatttgg ctcattttat atttagttcc ttccccagtc    940

aacaactgac cacatctgcc tctgcctgag cattaggatg ctcaaacatc ctatctttct   1000

tcttctattc atgcttttat ccattcttct ctgtcctgtc ttccctgctc caactctttc   1060

tctcaatatt cctgattttt ttttcaataa atttcacatg                         1100
```

<210> 44
<211> 265

<212> PRT
<213> Homo sapiens

<400> 44

Met Asn Arg Phe Leu Leu Leu Met Ser Leu Tyr Leu Leu Gly Ser Ala
1 5 10 15

Arg Gly Thr Ser Ser Gln Pro Asn Glu Leu Ser Gly Ser Ile Asp His
20 25 30

Gln Thr Ser Val Gln Gln Leu Pro Gly Glu Phe Phe Ser Leu Glu Asn
35 40 45

Pro Ser Asp Ala Glu Ala Leu Tyr Glu Thr Ser Ser Gly Leu Asn Thr
50 55 60

Leu Ser Glu His Gly Ser Ser Glu His Gly Ser Ser Lys His Thr Val
65 70 75 80

Ala Glu His Thr Ser Gly Glu His Ala Glu Ser Glu His Ala Ser Gly
85 90 95

Glu Pro Ala Ala Thr Glu His Ala Glu Gly Glu His Thr Val Gly Glu
100 105 110

Gln Pro Ser Gly Glu Gln Pro Ser Gly Glu His Leu Ser Gly Glu Gln
115 120 125

Pro Leu Ser Glu Leu Glu Ser Gly Glu Gln Pro Ser Asp Glu Gln Pro
130 135 140

Ser Gly Glu His Gly Ser Gly Glu Gln Pro Ser Gly Glu Gln Ala Ser
145 150 155 160

Gly Glu Gln Pro Ser Gly Glu His Ala Ser Gly Glu Gln Ala Ser Gly
165 170 175

Ala Pro Ile Ser Ser Thr Ser Thr Gly Thr Ile Leu Asn Cys Tyr Thr
180 185 190

Cys Ala Tyr Met Asn Asp Gln Gly Lys Cys Leu Arg Gly Glu Gly Thr
195 200 205

Cys Ile Thr Gln Asn Ser Gln Gln Cys Met Leu Lys Lys Ile Phe Glu
210 215 220

```
Gly Gly Lys Leu Gln Phe Met Val Gln Gly Cys Glu Asn Met Cys Pro
225                 230                 235                 240

Ser Met Asn Leu Phe Ser His Gly Thr Arg Met Gln Ile Ile Cys Cys
                245                 250                 255

Arg Asn Gln Ser Phe Cys Asn Lys Ile
            260                 265
```

<210> 45
<211> 1018
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (229)..(867)
<223>

<400> 45

```
gccaggcgaa ggcggagcgc taacgtctaa cgctaacggc ggtcgtgccc cgccgctgct     60

gtcacccccg gccgctgctg ccctccccgc cgaggttcta ctgctctcct tcttaagaag    120

ggtgggaggc actcggtctc tccccacacc tctcgcctga ggccaggcgc caggtgtcgc    180

ctgaagccag acagccggtt tgggagcgag cctgaggtca accaatca atg gct cag     237
                                                    Met Ala Gln
                                                    1

aca gat aag cca aca tgc atc ccg ccg gag ctg ccg aaa atg ctg aag      285
Thr Asp Lys Pro Thr Cys Ile Pro Pro Glu Leu Pro Lys Met Leu Lys
    5                   10                  15

gag ttt gcc aaa gcc gcc att cgg gcg cag ccg cag gac ctc atc cag      333
Glu Phe Ala Lys Ala Ala Ile Arg Ala Gln Pro Gln Asp Leu Ile Gln
20                  25                  30                  35

tgg ggg gcc gat tat ttt gag gcc ctg tcc cgt gga gag acg cct ccg      381
Trp Gly Ala Asp Tyr Phe Glu Ala Leu Ser Arg Gly Glu Thr Pro Pro
                40                  45                  50

gtg aga gag cgg tct gag cga gtc gct ttg tgt aac tgg gca gag cta      429
Val Arg Glu Arg Ser Glu Arg Val Ala Leu Cys Asn Trp Ala Glu Leu
            55                  60                  65

aca cct gag ctg tta aag atc ctg cat tct cag gtt gct ggc aga ctg      477
Thr Pro Glu Leu Leu Lys Ile Leu His Ser Gln Val Ala Gly Arg Leu
        70                  75                  80

atc atc cgt gca gag gag ctg gcc cag atg tgg aaa gtg gtg aat ctc      525
Ile Ile Arg Ala Glu Glu Leu Ala Gln Met Trp Lys Val Val Asn Leu
    85                  90                  95

cca aca gat ctg ttt aat agt gtg atg aat gtg ggt cgc ttc acg gag      573
Pro Thr Asp Leu Phe Asn Ser Val Met Asn Val Gly Arg Phe Thr Glu
```

```
100                 105                 110                 115

gag atc gag tgg ctg aag ttt tta gcc ctt gct tgc agc gct ctg gga      621
Glu Ile Glu Trp Leu Lys Phe Leu Ala Leu Ala Cys Ser Ala Leu Gly
                120                 125                 130

gtt act att acc aaa act ctc aag ata gtg tgt gag gtc tta tca tgt      669
Val Thr Ile Thr Lys Thr Leu Lys Ile Val Cys Glu Val Leu Ser Cys
            135                 140                 145

gac cac aat ggt ggg ttg ccc cga atc cca ttc agc acc ttc cag ttt      717
Asp His Asn Gly Gly Leu Pro Arg Ile Pro Phe Ser Thr Phe Gln Phe
        150                 155                 160

ctc tac acg tat att gcc gaa gtg gat ggg gag atc tgt gca tca cat      765
Leu Tyr Thr Tyr Ile Ala Glu Val Asp Gly Glu Ile Cys Ala Ser His
    165                 170                 175

gtc agc agg atg cta aac tac att gaa cag gaa gta att ggt cct gat      813
Val Ser Arg Met Leu Asn Tyr Ile Glu Gln Glu Val Ile Gly Pro Asp
180                 185                 190                 195

ggt tta atc acg gtg aat gac ttt acc caa aac ccc agg gtt tgg ctg      861
Gly Leu Ile Thr Val Asn Asp Phe Thr Gln Asn Pro Arg Val Trp Leu
            200                 205                 210

gag taa cagcacaatt ttggcaattt taaaggaaga tacagaggtg attgtacttc      917
Glu

agaatgataa acccatatac cacctaaaat caatttttctt gtacaactgg tacacactaa    977

taaacaaaca tgtgagatca gaaaaaaaaa aaaaaaaaaa a                        1018

<210>  46
<211>  212
<212>  PRT
<213>  Homo sapiens

<400>  46

Met Ala Gln Thr Asp Lys Pro Thr Cys Ile Pro Pro Glu Leu Pro Lys
1               5                   10                  15

Met Leu Lys Glu Phe Ala Lys Ala Ala Ile Arg Ala Gln Pro Gln Asp
            20                  25                  30

Leu Ile Gln Trp Gly Ala Asp Tyr Phe Glu Ala Leu Ser Arg Gly Glu
        35                  40                  45

Thr Pro Pro Val Arg Glu Arg Ser Glu Arg Val Ala Leu Cys Asn Trp
    50                  55                  60

Ala Glu Leu Thr Pro Glu Leu Leu Lys Ile Leu His Ser Gln Val Ala
65                  70                  75                  80
```

Gly Arg Leu Ile Ile Arg Ala Glu Glu Leu Ala Gln Met Trp Lys Val
85 90 95

Val Asn Leu Pro Thr Asp Leu Phe Asn Ser Val Met Asn Val Gly Arg
100 105 110

Phe Thr Glu Glu Ile Glu Trp Leu Lys Phe Leu Ala Leu Ala Cys Ser
115 120 125

Ala Leu Gly Val Thr Ile Thr Lys Thr Leu Lys Ile Val Cys Glu Val
130 135 140

Leu Ser Cys Asp His Asn Gly Gly Leu Pro Arg Ile Pro Phe Ser Thr
145 150 155 160

Phe Gln Phe Leu Tyr Thr Tyr Ile Ala Glu Val Asp Gly Glu Ile Cys
165 170 175

Ala Ser His Val Ser Arg Met Leu Asn Tyr Ile Glu Gln Glu Val Ile
180 185 190

Gly Pro Asp Gly Leu Ile Thr Val Asn Asp Phe Thr Gln Asn Pro Arg
195 200 205

Val Trp Leu Glu
210

<210> 47
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 47
gcaatggctg gaggagaact 20

<210> 48
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 48
agccactttt agccacttca tc 22

<210> 49
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 49
tgtgtgactc catcctctac 20

<210> 50
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 50
gtctggcttt ttgtgtgtgt g 21

<210> 51
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 51
gaagacacgg aaggcacaga 20

<210> 52
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 52
agccactttt agccactcat c 21

<210> 53
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 53
accggaaact catcacccca at 22

<210> 54
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 54
gtaagcaaag ccaggaaagt g 21

<210> 55
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 55
gcaatggctg gaggagaact 20

<210> 56
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 56
taaactggta tcctgtgtgt ga 22

<210> 57
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 57
tgtgtgactc catcctctac 20

<210> 58
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 58
aggtagagca cgtagtcatc 20

<210> 59
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 59
cccgagtctt ctggtggtta 20

<210> 60
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 60
agcattgaca ggttgggtat c 21

<210> 61
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 61
ggccacgcgt cgactagtac 20

<210> 62
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 62
agttctcctc cagccat 17